

STUDY OF PEPTIDE INTERACTIONS IN SOLUTION THROUGH THE USE OF LOCAL CORRELATION METHODS

Dissertation
for the award of the degree
“Doctor rerum naturalium” (Dr.rer.nat)
of the Georg-August-Universität Göttingen

within the doctoral program of Chemistry
of the Georg-August University School of Science (GAUSS)

submitted by
João Carlos Agostinho de Oliveira
from Caldas da Rainha, Portugal

Göttingen, 2014

Thesis Committee

Jun.-Prof. Dr. Ricardo Mata

Institute for Physical Chemistry, Georg-August-Universität Göttingen

Prof. Dr. Peter Botschwina

Institute for Physical Chemistry, Georg-August-Universität Göttingen

Dr. Nuno Galamba

Ecole Normale Supérieure-PSL Research University, Département de Chimie, Sorbonne
Universités - UPMC Univ Paris, CNRS UMR 8640 PASTEUR

Members of the Examination Board

Reviewer:

Jun.-Prof. Dr. Ricardo Mata

Institute for Physical Chemistry, Georg-August-Universität Göttingen

Second Reviewer:

Prof. Dr. Peter Botschwina

Institute for Physical Chemistry, Georg-August-Universität Göttingen

Further members of the Examination Board

Prof. Dr. Martin Suhm

Institute for Physical Chemistry, Georg-August-Universität Göttingen

Prof. Dr. Dietmar Stalke

Institute for Inorganic Chemistry, Georg-August-Universität Göttingen

Prof. Dr. Burkhard Geil

Institute for Physical Chemistry, Georg-August-Universität Göttingen

Dr. Jochen Hub

Department of Molecular Structural Biology, Georg-August-Universität Göttingen

Date of the oral examination: 14.08.2014

Declaration

I hereby declare that I have created this thesis without the help of third parties and by my own efforts. No other than the referenced sources have been employed and knowledge or ideas taken directly or indirectly from prior existing works were referenced as such.

João Carlos Agostinho de Oliveira

Contents

Acknowledgements	VII
Abstract	IX
Abbreviations	XI
1 Introduction	1
2 Theoretical background	5
2.1 Quantum Mechanics	5
2.1.1 Hartree-Fock	8
2.1.2 Møller-Plesset Perturbation Theory	11
2.1.3 Local Correlation Methods	15
2.1.4 Density Functional Theory	20
2.2 Molecular Mechanics	23
2.2.1 Metropolis Monte Carlo	25
2.3 Quantum Mechanics/Molecular Mechanics	29
2.4 Implicit Solvation Models	30
3 Local MP2 in combination with COSMO	35
3.1 Method	35
3.2 Test systems	38
3.2.1 Computational Details	38
3.2.2 Benchmark Results	39
3.2.3 Amino acid Interactions	42
3.2.4 β -Peptide Conformers	46

4	Specific ion-amino acid interactions	51
4.1	Overview	51
4.2	Anion Specific Effects	52
4.2.1	Computational details	53
4.2.2	Results and Discussion	54
5	β-Peptides conformational studies	69
5.1	Overview	69
5.2	Computational Details	72
5.3	Results and Discussion	75
5.3.1	β -peptides with cyclic side chains	75
5.3.2	β -peptides with aliphatic side chains	84
6	Perturbative-Monte Carlo QM/MM	97
6.1	Method	97
6.1.1	Water model	100
6.1.2	Translation and rotation movements	100
6.1.3	Simulation of non-periodic systems	102
6.2	Test systems	103
7	Conclusions	111
	Bibliography	117
	Curriculum Vitae	129

Acknowledgements

First of all, I would like to thank Prof. Dr. Mata for welcoming me and for giving me the opportunity to develop my research in his research group. I am also very thankful for his encouragement, guidance and patience; without him this thesis would not have been possible.

My sincere thanks to Prof. Dr. Peter Botschwina for the nice welcoming in the Institute of Physical Chemistry.

I would like to express my gratitude to Dr. Galamba, who together with Prof. Mata introduced me to the science behind the stability of proteins. Also to Dr. Ramana and Prof. Dr. Diederichsen for presenting me the challenging world of β -peptides, and for the pleasant collaborations.

I gratefully thank Dr. Oswald for his assistance in preparing the PES plots.

I acknowledge the financial support from the German Excellence Initiative, through the Free Floater Research Group program of the University of Göttingen.

My thanks to the past and actual members of the Computational Chemistry and Biochemistry Group, for their help in numerous situations and for creating a pleasant and enthusiastic environment for everyday work: Cecilia Rosso, Johannes Dieterich, Jonas Feldt, Thorsten Stolper, Milica Andrejić, Diego Andrada, Martin Werner, Thorsten Teuteberg and Axel Wuttke.

Heartfelt thanks to my friend Catarina for all the laughs and the great moments that we spent together. For helping me bear the tough moments, to her a Muito Obrigado!

I also would like to thank my friends Margarida and Wolfgang for their unconditional support and strength to keep on going.

Last but not least, I would like to thank my parents, sister and brother, and remaining family for their unconditional support, to whom I will dedicate some words in portuguese:

Gostaria, por fim, de agradecer aos meus pais, irmã e irmão e restante família pela força, motivação e apoio que me deram ao longo dos tempos. Sem as últimas nunca teria alcançado o que alcancei. Estou-lhes eternamente grato.

Abstract

The forces that control peptides conformations in solution have intrigued scientists over the decades. Understanding the former is of high interest, since they present a direct impact in the function and stability of enzymes and proteins.

The emphasis of this work is on the study of peptide-ion interactions, as well as the effect of the solvent in the overall peptide conformation. The latter was explored for the specific case of β -peptides. These are known to present an intrinsic propensity to fold into a variety of secondary structures. These features were verified in a variety of solvent environments and peptidic chains.

In the present thesis, one proposes to study such effects by quantum mechanics. From the several methods available, one gives focus to local correlation methods, namely LMP2. Solvent effects were introduced by using an implicit solvation model. For the latter, COSMO was the model of choice. The COSMO-LMP2 model was shown to be quite promising in the treatment of amino acid-anions specific interactions, as well as in the study of the stability of β -peptides conformers. The final work is dedicated to the development of a Perturbative Monte Carlo QM/MM scheme. Under this approach, the MM steps are treated perturbatively in their interaction with the QM system. The results revealed the potential of such approach in the treatment of properties of complex systems.

Abbreviations

BSSE	Basis set superposition error
<i>cis</i> -ACPC	<i>cis</i> -2-aminocyclopentanecarboxylic acid
<i>cis</i> -FAA	<i>cis</i> -3-aminotetrahydrofuran-2-carboxylic acid
COSMO	Conductor-like screening model
CSM	Continuum solvation model
DF	Density-fitting
DFT	Density functional theory
FF	Force field
GGA	Generalized gradient approximation
HF	Hartree-Fock approximation
LDA	Local density approximation
LMO	Localized molecular orbital
LMP2	Local Møller-Plesset perturbation theory of second order
MC	Monte Carlo
MD	Molecular dynamics
MM	Molecular mechanics
MP2	Møller-Plesset perturbation theory of second order
OPLS	Optimized Potentials for Liquid Simulations
PCM	Polarizable continuum model
PertMC	Perturbative Monte Carlo
PES	Potential energy surface
QM	Quantum Mechanics
QM/MM	Quantum Mechanics/Molecular Mechanics
SASA	Solvent-accessible surface area
SCF	Self-consistent field method

1 Introduction

The study of the physical chemical processes that control the conformational behaviour of peptides in solution is the centrepiece of this work. It has been the subject of research for generations of scientists, as we try to better understand the functionality and stability of proteins and enzymes.

In the later years of the nineteen century, Franz Hofmeister explored the stability of proteins in ionic aqueous solution. The experiments involved the addition of several salts to a solution of egg-white protein. He verified that some salts lead to the precipitation of the protein (salting-out) while others maintained the protein in solution (salting-in). Following his observations, he ordered the ions in a series which bear his name, the Hofmeister series. This discovery had a significant impact towards proteins purification, and is still used in our days for such purpose. However, the effects behind such processes are not completely understood. They were originally explained in terms of the ion effects in the hydrogen bond network of liquid water. However, recent experimental findings contradict such theory.¹⁻³ They lead one to believe that ion-water interactions do not play a major role. Therefore, one needs to take into consideration other effects, such as ion-protein and protein-water interactions. Part of the present manuscript is dedicated to explore ion-protein specific interactions in solution.

Natural proteins are predominantly built from α -amino acids. Closely related are the β -amino acids, which are mostly synthetic. These present an extra carbon between the amine and the carboxyl groups. This allows them to fold into several secondary structures, such as helices and sheet like conformations, separated by only a few $\text{kcal}\cdot\text{mol}^{-1}$ in the potential energy surface with high dependence on the solvent. Contrary to α -peptides, they present high resistance to enzymatic degradation. In this way, although unnatural, with the exception of β -alanine, this family of peptides present a high potential in drug design. In this work one makes use of the high variety of secondary structures adopted by β -peptides to better understand the solution impact on the interactions between the peptidic chains.

In computational studies, the effect of the solvent can be simulated explicitly or implicitly. In the first case one makes use of an atomistic description of the solvent, while in the latter the solvent is represented by a dielectric continuum.^{4,5} The use of explicit solvent models is commonly found in Molecular Dynamics and Monte Carlo simulations. The main disadvantage of such approaches is the need of a high number of solvent configurations in order to obtain a reasonable description. Models that use an implicit approach are known as continuum solvation models (CSMs).^{4,5} Among the various CSMs developed to date, the conductor-like screening model (COSMO)^{6,7} is one of the most popular approaches. It holds a low computational cost and has proven to deliver good accuracy in the treatment of polar solvents. Alternative formulations, such as the polarizable continuum model (PCM),^{8,9} may hold more developed approximations for the external potential or even include empirical corrections for dispersion.¹⁰ These may prove important when dealing with apolar solvents, where the latter forces have a stronger weight in the overall interaction. However, the agreement between different CSMs is fairly good, and for most applications the computed results will not depend strongly on the choice made.

When compared with an explicit solvation model the use of CSM avoids the need of an extensive search of solvent configurations treated with statistical mechanics. This results in significant computational time savings. In this way the use of CSM is quite advantageous, but in circumstances where specific interactions between the solute and the solvent are important an explicit description of the latter is necessary. This is considered to be the main disadvantage of the CSMs. A possible way to overcome this is to introduce some explicit solvent molecules in areas where such interactions take place, and simulate the effect of the rest of the solvent by a CSM. As an alternative one can make use of hybrid QM/MM methods.

In QM/MM schemes the solvent is treated explicitly, while a specific solute can be treated by quantum mechanics.^{11–14} Assuming a solvated biomolecule, the latter could be treated at the QM level, while the solvent is treated at the MM level through the use of a force field. Time development, however, requires that both solute and solvent energies/gradients are computed at each step. This leads to a high computational cost since the number of degrees of freedom to be sampled in solution is extremely high. The latter could be alleviated if one could decouple the QM from the MM movements. Such an approach was proposed by Truong *et al.* in the development of a perturbative Monte Carlo QM/MM scheme.^{15,16} The main advantage of such method is the fact that

it avoids the need to perform a full QM calculation per every solvent movement. This method has been explored in this work in an attempt to bridge explicit and implicit solvation models.

In the quantum mechanical methods the main challenges, besides the choice between the several approaches available, is the scaling of the computational time with the system size. This mainly results from the use of canonical orbitals as for example in Hartree-Fock and post-Hartree-Fock methods, the latter used in the treatment of electronic correlation. A way to overcome such disadvantage was proposed by Pulay and Saebø in the late years of the past century.¹⁷⁻²¹ Taking into account the local character of dynamic correlation, the authors proposed to treat the correlation through local orbital spaces. This led to a large family of local correlation methods. The latter became over the last years an established approach for the study of ground state properties, with extensions to excited states already developed.²²⁻²⁴ Not only do they accomplish drastic savings in the computational cost, they also reduce the basis set superposition error by construction. Such properties make them state-of-the-art tools in the study of weakly interacting systems. Although it would be of great interest to use such methods in the treatment of condensed systems, applications have been mostly restricted to gas phase molecules. In this way it would be advantageous to couple such approaches with CSMs.

One starts in Chapter 3 by discussing an implementation of COSMO-LMP2 analytical gradients and its use in the optimization of molecular structures in solution. A benchmark study is presented involving several model systems. The geometries obtained at the COSMO-LMP2 level of theory were characterized and compared with the canonical approach. An extension to the evaluation of the interaction energy between an amino acid model system, halide anions and water is also presented for gas phase and solution. At the end of the chapter one presents a complementary study that involves the application of COSMO-LMP2 in the study of the conformational stability of a β -peptide tetramer.

In Chapter 4 one probes the interaction of specific anions with several amino acids in solution through the use of COSMO-LMP2. In order to address the ionic affinity toward different amino acid residues, one focuses on anions with the same valence, specifically halide anions. The latter are positioned in the Hofmeister series in the relative order $[F^- > Cl^- > Br^- > I^-]$, where I^- is a salting-in anion, and at the left end, F^- (salting-out). A detailed view of the interaction energy of such systems is presented. Based on the latter the ordering of the anions in the mentioned series is

discussed and their effects towards protein stability in ionic solutions.

Chapter 5 is dedicated to the study of the conformational landscape from β -peptides in solution. The latter are known to present a more flexible backbone than α -peptides. In this section one studies the effect of the side chain, as well as solvent environment on the folding of such peptides. For this purpose several hexapeptides composed of different β -amino acids were investigated. The latter are separated into two groups, amino acids with cyclic and aliphatic side chains. The conformations were analysed in gas phase and in solution, where the effects of the latter were introduced by the COSMO model.

Chapter 6 is dedicated to the implementation of a Perturbative Monte Carlo QM/MM based on the seminal work of Truong *et al.*^{15,16} This approach is based on the assumption that a movement of a molecule in the MM region does not affect significantly the wavefunction from the QM molecule. This allows one to treat the MM movements perturbatively. This procedure avoids the need to perform a full QM calculation for each MM step, the first is only executed after a certain number of perturbative steps in order to update the wavefunction of the QM part. One presents details of such perturbative approach as well as some preliminary results.

2 Theoretical background

2.1 Quantum Mechanics

In Quantum Mechanics, all the system information can be found in its wavefunction, Ψ . In order to access a certain property, one needs to apply the appropriate operator to the wavefunction. An example is the Schrödinger equation. In the latter the energy of a certain state is obtained by applying the Hamiltonian operator to the wavefunction.

For a system composed of M nuclei and N electrons in the absence of an external potential the Hamiltonian in atomic units can be described as

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 - \frac{1}{2} \sum_{A=1}^M \frac{\nabla_A^2}{M_A} - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{r_{AB}}. \quad (2.1)$$

In the former the indices i and j are related to the electrons, while the indices A and B run over the nuclei. M and Z represent the masses and atomic charges, respectively, and r is the interatomic or interelectron distances.

Using the kinetic and potential operator notation to describe each term of equation (2.1) one obtains

$$\hat{H} = \hat{T}_e + \hat{T}_n + \hat{V}_{ne} + \hat{V}_{ee} + \hat{V}_{nn}. \quad (2.2)$$

In the previous equations (2.1 and 2.2) the first two terms take into account the electronic (e) and nuclear (n) kinetic energy, while the remaining are related to Coulombic interactions. The latter include the attraction between electron and nuclei, the repulsion between electrons and repulsion between nuclei.

The analytical resolution of the Schrödinger equation can not be obtained for a many-electron system. This is mainly related to the fact that the eigenfunctions of the Hamiltonian, in equation (2.1), depend explicitly on the coordinates of all electrons and nuclei of the system in study. In order to overcome this problem, one can make use of a simplification proposed by Julius Oppenheimer and Max Born,²⁵ which allows one to treat

the motion of electrons and nuclei separately.

The Born-Oppenheimer approximation is based on the fact that the nuclei are much heavier than the electrons. When compared to the nuclei the electrons move at a much higher speed. In this way the nuclei can be considered as fixed when compared to the electrons and the Coulombic repulsion between the nuclei considered as constant.

This approximation simplifies the eigenvalue problem, since the electronic wavefunction is parametrized in relation to the nuclei depending only on the coordinates of the electrons. This although considered a mild approximation, leads to the fundamental concept of a potential energy surface (PES), where the latter is defined as the electronic energy over all possible nuclear coordinates.²⁶ Such approximation breaks down when two solutions (states) of the electronic Schrödinger equation come close together energetically, as in the case of conical intersections.²⁷

A consequence of this approximation the total energy of the system can be defined as

$$E_{tot} = E_{elec} + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{r_{AB}}, \quad (2.3)$$

where the electronic energy, E_{elec} , can be obtained by solving the time-independent Schrödinger equation only for the electronic degrees of freedom. One can then rewrite the Schrödinger equation as

$$\hat{H}_{elec} \Psi_{elec} = E_{elec} \Psi_{elec}. \quad (2.4)$$

where Ψ_{elec} is the wavefunction for the N electrons present in the system.

Since one is focused in discussing the electronic eigenvalue problem one will omit from now on the *elec* subscript in equation (2.4). The Coulombic repulsion between nuclei (\hat{V}_{nn}) will also be omitted from now on, but the addition of the latter can be made a posteriori in order to obtain the total energy of the system. In the previous equation the Hamiltonian operator is Hermitian, the eigenvalues are real and its eigenfunctions are orthogonal, normalized and form a complete set.²⁷ The wavefunction must present the following properties: it needs to be continuous, single-valued, square integrable, must not be zero everywhere (if this was true the system just does not exist) and must vanish to infinity.²⁸

Let us assume that one picks a trial wave function, Φ , that obeys to the above mentioned constraints and that is operated by the Hamiltonian. If $\{\Psi_i\}$ is the complete set of orthonormal wavefunctions, one can write Φ as a linear combination of Ψ_i , i.e.

$$\Phi = \sum_i^{\infty} c_i \Psi_i, \quad (2.5)$$

where c_i are the coefficients on the basis formed by the eigenfunctions. The energy associated to this trial wavefunction, E_{trial} can be calculated as

$$\begin{aligned} E_{trial} &= \langle \Phi | \hat{H} | \Phi \rangle \\ &= \sum_i^{\infty} \sum_j^{\infty} \langle \Phi | \Psi_i \rangle \langle \Psi_i | \hat{H} | \Psi_j \rangle \langle \Psi_j | \Phi \rangle \\ &= \sum_i^{\infty} \sum_j^{\infty} \langle \Phi | \Psi_i \rangle \delta_{ij} E_j \langle \Psi_j | \Phi \rangle \\ &= \sum_i^{\infty} \langle \Phi | \Psi_i \rangle E_i \langle \Psi_i | \Phi \rangle \\ &= \sum_i^{\infty} c_i^2 E_i, \end{aligned} \quad (2.6)$$

where δ_{ij} is the Kronecker delta, whose value is 1 when $i = j$ otherwise is 0. One can see that the trial energy is defined as the sum of the eigenvalues of \hat{H} weighed by the coefficients c_i .²⁶ In the set of all possible E_i there must be one that presents the lowest energy possible. The latter is the ground state energy (E_0). Therefore the trial energy must be larger or equal to the ground state energy ($E_{trial} \geq E_0$). By definition, $(E_i - E_0)$ must be greater than or equal to zero,²⁶ so that

$$\begin{aligned} \langle \Phi | \hat{H} | \Phi \rangle - E_0 \langle \Phi | \Phi \rangle &= \sum_i^{\infty} c_i^2 (E_i - E_0) \\ \langle \Phi | \hat{H} | \Phi \rangle - E_0 \langle \Phi | \Phi \rangle &\geq 0 \\ \frac{\langle \Phi | \hat{H} | \Phi \rangle}{\langle \Phi | \Phi \rangle} &\geq E_0. \end{aligned} \quad (2.7)$$

A consequence of the Variational Principle is that any energy obtained with the trial wavefunction will always be higher than the exact energy of the system. Therefore the solution of the Schrödinger equation can be formulated as a minimization problem.

2.1.1 Hartree-Fock

In order to solve the Schrödinger equation to a system of N -electron, one can make use from the Hartree-Fock (HF) approximation. In the latter the wavefunction of a system is represented by a Slater determinant as

$$\Psi^{HF} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(\mathbf{x}_1) & \psi_2(\mathbf{x}_1) & \cdots & \psi_N(\mathbf{x}_1) \\ \psi_1(\mathbf{x}_2) & \psi_2(\mathbf{x}_2) & \cdots & \psi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(\mathbf{x}_N) & \psi_2(\mathbf{x}_N) & \cdots & \psi_N(\mathbf{x}_N) \end{vmatrix} = \Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N), \quad (2.8)$$

where N is the normalization constant, ψ are molecular spin orbitals and \mathbf{x} is the vector with both spatial and spin coordinates from the electron i , $\mathbf{x}_i = \{\mathbf{r}_i, \mathbf{s}_i\}$. For a closed shell configuration, one can use the same spatial orbitals, ϕ , for both α and β spin electrons.

A Slater determinant is only exact for non-interacting systems. The former is known to assure the antisymmetry of the wavefunction when applied to a system of N -electrons. As a first approximation the HF takes advantage of the last property and applies it to an interacting system. This simplification is the basis for general wavefunction theory. Solving the Schrödinger equation under this ansatz means finding a set of spin orbitals that build the single Slater determinant with the lowest energy.

According to the Slater-Condon rules²⁹ when the Hamiltonian operates on a Slater determinant it generates a sum of one- and two-electron contributions.²⁹ So when this operates on the determinant (2.8) for the closed shell system taking already into account the spin integration leads to²⁹

$$E^{HF} = \langle \Psi^{HF} | \hat{H} | \Psi^{HF} \rangle = 2 \sum_i^{N/2} (i | \hat{h} | i) + \sum_i^{N/2} \sum_j^{N/2} 2(i | j j) - (i j | j i), \quad (2.9)$$

where E^{HF} is the Hartree-Fock energy. The one- and two-electron integrals are defined as

$$(i | \hat{h} | i) = \int \phi_i^*(\mathbf{r}_1) \hat{h}(\mathbf{r}_1) \phi_i(\mathbf{r}_1) d\mathbf{r}_1 \quad (2.10)$$

and

$$\begin{aligned}
 (ii|jj) &= \iint \phi_i^*(\mathbf{r}_1) \phi_j^*(\mathbf{r}_2) r_{12}^{-1} \phi_i(\mathbf{r}_1) \phi_j(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \\
 &= J_{ij} \\
 (ij|ji) &= \iint \phi_i^*(\mathbf{r}_1) \phi_j^*(\mathbf{r}_2) r_{12}^{-1} \phi_j(\mathbf{r}_1) \phi_i(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \\
 &= K_{ij}
 \end{aligned} \tag{2.11}$$

respectively. In the previous equations J_{ij} is the Coulomb integral which represents the classical repulsion between two electrons occupying two spatial orbitals i and j , while K_{ij} is the non-classical exchange integral. In the case of $i = j$ both integrals cancel out since an electron does not interact with itself.

The set of orbitals that minimize the HF energy, E^{HF} , can be found using the Lagrangian function

$$\mathcal{L} = E^{HF} - 2 \sum_{ij}^{N/2} l_{ij} (\langle i|j \rangle - \delta_{ij}), \tag{2.12}$$

where E^{HF} is defined as in equation (2.9), the l_{ij} are the Lagrange multipliers and $\langle i|j \rangle$ is an overlap integral. Since \mathcal{L} is real and $\langle i|j \rangle = (j|i)^*$ then $l_{ij} = l_{ji}^*$. Under an orthonormality constraint, $\langle i|j \rangle = \delta_{ij}$, the minimal energy can be obtained by setting the Lagrange derivative to zero in respect to orbital i . This leads to a set of coupled equations

$$\hat{f} |i\rangle = \sum_j^{N/2} l_{ij} |j\rangle, \tag{2.13}$$

where \hat{f} is the Fock operator. The latter is defined as

$$\begin{aligned}
 \hat{f}(i) &= \hat{h}(i) + \sum_j^{N/2} [2J_j - K_j] \\
 &= \hat{h}(i) + \hat{g}(i)
 \end{aligned} \tag{2.14}$$

where $\hat{h}(i)$ is a one-electron operator and $\hat{g}(i)$ is a two-electron operator. The latter represents the mean field potential generated by the electrons. J and K are the Coulomb

and exchange operators. They are best defined when operating in a spatial orbital ϕ_i as

$$\hat{J}_j(1)\phi_i(\mathbf{r}_1) = \int \phi_j^*(\mathbf{r}_2)r_{12}^{-1}\phi_j(\mathbf{r}_2)\phi_i(\mathbf{r}_1)d\mathbf{r}_2 \quad (2.15)$$

$$\hat{K}_j(1)\phi_i(\mathbf{r}_1) = \int \phi_j^*(\mathbf{r}_2)r_{12}^{-1}\phi_j(\mathbf{r}_1)\phi_i(\mathbf{r}_2)d\mathbf{r}_2. \quad (2.16)$$

Any single determinant wavefunction contains a certain flexibility and the spin orbitals can be mixed without changing the eigenvalue. In order to obtain HF in the canonical form one needs to diagonalize the Fock operator (equation (2.13)) using an unitary transformation of the spin orbitals.²⁹ This gives origin to the so called HF canonical equations:

$$\hat{f}|i\rangle = \varepsilon_i|i\rangle, \quad (2.17)$$

where eigenvalues ε_i are the orbital energies.

The Hartree-Fock approximation simplifies the N -particle problem to a set of coupled equations. Under this approximation each electron only feels the average potential produced by all the other electrons. These are in fact a set of pseudo-eigenvalue equations since the Fock operator depends on all occupied molecular orbitals.²⁷ One Fock orbital can only be determined if all the occupied molecular orbitals are known, therefore the problem needs to be solved iteratively.²⁷ In order to solve these equations one question still remains, the definition of the spatial orbitals.

For a molecular system the spatial molecular orbitals are described as a linear combination of atomic orbitals (LCAO),

$$\phi_i(\mathbf{r}) = \sum_{\mu} C_{i\mu}\chi_{\mu}(\mathbf{r}), \quad (2.18)$$

where χ_{μ} is an atomic orbital and $C_{i\mu}$ is the coefficient matrix. The Hartree-Fock energy in equation (2.9) can be written in the atomic orbital (AO) basis as

$$\begin{aligned} E^{HF} &= \sum_{\mu\nu} P_{\mu\nu}h_{\mu\nu} + \frac{1}{2} \sum_{\mu\nu} \sum_{\rho\sigma} P_{\mu\nu}P_{\rho\sigma} \left\{ (\mu\nu|\rho\sigma) - \frac{1}{2}(\mu\sigma|\rho\nu) \right\} \\ &= \frac{1}{2} \sum_{\mu\nu} P_{\mu\nu}(h_{\mu\nu} + f_{\mu\nu}), \end{aligned} \quad (2.19)$$

where $h_{\mu\nu}$ and $f_{\mu\nu}$ are integral matrices for the operators \hat{h} and \hat{f} in the atomic orbital

basis. The matrix P is the one-electron density matrix defined as

$$P_{\mu\nu} = 2 \sum_i^{N/2} C_{\mu i} C_{\nu i}^*. \quad (2.20)$$

The computational bottleneck in the HF procedure arises from the two-electron integrals, defined as

$$(\mu\nu|\lambda\sigma) = \int \chi_\mu^*(i) \chi_\nu(i) r_{ij}^{-1} \chi_\lambda^*(j) \chi_\sigma(j) d\mathbf{r}_i d\mathbf{r}_j \quad (2.21)$$

that are in great number.²⁹ The calculation and transformation of these four-index two electron integrals lead to a scaling $\mathcal{O}(\mathcal{N}^4)$ where \mathcal{N} is a measure of the system size.²⁶ Hartree-Fock is able to retrieve as much as 99 % of the total energy for simple organic molecules.³⁰ The remaining energy is due to instantaneous interactions between the electrons, designated as electron correlation. The latter is indispensable when describing reactions where bond breaking and bond formation are involved.³⁰ The electron correlation energy can then be defined as the difference between the HF energy and the total non-relativistic energy of the system.³⁰ In order to retrieve this energy one can make use of so called post-Hartree-Fock methods. The choice of the latter can then be a compromise of accuracy and computational time, since not all are feasible due to the dimensions of the system. Perhaps the most commonly used correlation methods are based on Møller-Plesset Perturbation Theory. This will be discussed in the following section.

2.1.2 Møller-Plesset Perturbation Theory

Møller-Plesset Perturbation Theory takes advantage from the fact that HF theory is able to describe quite well the energy of the system and introduces into the Hamiltonian an additional term in order to retrieve the dynamic correlation. In this way the Hamiltonian is composed of two parts, one reference $\hat{H}^{(0)}$ and one perturbation $\hat{H}^{(1)}$. Introducing this approach in the Schrödinger equation, this will take the form

$$(\hat{H}^{(0)} + \hat{H}^{(1)}) |\Psi\rangle = E |\Psi\rangle, \quad (2.22)$$

with

$$\begin{aligned}\hat{H}^{(0)} &= \sum_i^N \hat{f}(i) \\ &= \sum_i^N [\hat{h}(i) + \hat{g}(i)]\end{aligned}\tag{2.23}$$

$$\hat{H}^{(1)} = \hat{H} - \hat{H}^{(0)}.$$

The choice of the unperturbed Hamiltonian in the form of a sum over Fock operators leads to the so called Møller-Plesset (MP k) perturbation theory, where k is a cardinal number that indicates the order of the perturbation. According to Rayleigh-Schrödinger perturbation theory one can estimate the eigenfunctions and eigenvalues of a more complex operator, taking advantage of a simpler one.²⁷ This is done by expanding the eigenvalues and eigenfunction in a Taylor series as²⁶

$$\begin{aligned}E &= \sum_{k=0}^N \lambda^k E^{(k)} = E^{(0)} + \lambda^1 E^{(1)} + \lambda^2 E^{(2)} + \dots \\ \Psi &= \sum_{k=0}^N \lambda^k \Psi^{(k)} = \Psi^{(0)} + \lambda^1 \Psi^{(1)} + \lambda^2 \Psi^{(2)} + \dots,\end{aligned}\tag{2.24}$$

where the superscripts k are the order of correction to the zeroth order. Inserting equation (2.24) in (2.22), one obtains a set of equations

$$\begin{aligned}\lambda^0 : \hat{H}^{(0)} |\Psi^{(0)}\rangle &= E^{(0)} |\Psi^{(0)}\rangle \\ \lambda^1 : \hat{H}^{(0)} |\Psi^{(1)}\rangle + \hat{H}^{(1)} |\Psi^{(0)}\rangle &= E^{(0)} |\Psi^{(1)}\rangle + E^{(1)} |\Psi^{(0)}\rangle \\ \lambda^2 : \hat{H}^{(0)} |\Psi^{(2)}\rangle + \hat{H}^{(1)} |\Psi^{(1)}\rangle &= E^{(0)} |\Psi^{(2)}\rangle + E^{(1)} |\Psi^{(1)}\rangle + E^{(2)} |\Psi^{(0)}\rangle \\ \dots & \\ \lambda^k : \hat{H}^{(0)} |\Psi^{(k)}\rangle + \hat{H}^{(1)} |\Psi^{(k-1)}\rangle &= \sum_{i=0}^n E^{(i)} |\Psi^{(k-i)}\rangle.\end{aligned}\tag{2.25}$$

MP takes as a reference the HF wavefunction $\Psi^{(0)} = \Psi^{HF}$. In order to obtain the correlation energy one needs to multiply equation (2.25) on the left by $|\Psi^{(0)}\rangle$ which

leads, in the case of second order, to

$$\begin{aligned}\lambda^0 : E^{(0)} &= \langle \Psi^{(0)} | \hat{H}^{(0)} | \Psi^{(0)} \rangle \\ \lambda^1 : E^{(1)} &= \langle \Psi^{(0)} | \hat{H}^{(1)} | \Psi^{(0)} \rangle \\ \lambda^2 : E^{(2)} &= \langle \Psi^{(0)} | \hat{H}^{(1)} | \Psi^{(1)} \rangle.\end{aligned}\tag{2.26}$$

From the previous equations one can see that the sum of the two first order terms (0, 1) give the HF energy, $E^{HF} = E^{(0)} + E^{(1)}$. The first correction to the reference energy will be obtained at second order. This requires a first order wavefunction ($\Psi^{(1)}$). According to the Brillouin's theorem single excited determinants do not interact with the reference determinant.²⁷ In this way $\Psi^{(1)}$ is expanded in doubly excited determinants, Φ_{ij}^{ab} .³¹ The latter is defined with the help of second quantization³¹ as

$$|\Phi_{ij}^{ab}\rangle = \hat{E}_{ai}\hat{E}_{bj} |\Psi^{HF}\rangle, \tag{2.27}$$

where the \hat{E}_{mn} is a single excitation operator that excites the electron from the occupied orbital n to the virtual orbital m . The first order wavefunction $\Psi^{(1)}$ is then constructed as a linear combination of different excited configurations³¹

$$\Psi^{(1)} = \frac{1}{2} \sum_{ij} \sum_{ab} T_{ab}^{ij} \Phi_{ij}^{ab}, \tag{2.28}$$

where T_{ab}^{ij} are the amplitudes. The single excitation operators in equation (2.27) commute and therefore $\Phi_{ij}^{ab} = \Phi_{ji}^{ba}$ and $T_{ab}^{ij} = T_{ba}^{ji}$. The definition used for the double excited configuration in equation (2.27) presents the disadvantage that these are not orthonormal. The problem can be alleviated through the use of contravariant configurations $\tilde{\Phi}_{ij}^{ab}$ and amplitudes \tilde{T}_{ab}^{ij} ,³¹

$$\begin{aligned}\tilde{\Phi}_{ij}^{ab} &= \frac{1}{6} (2\Phi_{ij}^{ab} + \Phi_{ji}^{ab}) \\ \tilde{T}_{ab}^{ij} &= 2T_{ab}^{ij} - T_{ab}^{ji}.\end{aligned}\tag{2.29}$$

Equation (2.28) can then be rewritten as

$$\Psi^{(1)} = \frac{1}{2} \sum_{ij} \sum_{ab} T_{ab}^{ij} \Phi_{ij}^{ab} = \sum_{ij} \sum_{ab} \tilde{T}_{ab}^{ij} \tilde{\Phi}_{ij}^{ab}. \tag{2.30}$$

The contravariant configurations have the following properties:

$$\begin{aligned}
\langle \tilde{\Phi}_{ij}^{ab} | \Phi_{kl}^{cd} \rangle &= \delta_{ac} \delta_{bd} \delta_{ik} \delta_{jl} + \delta_{ad} \delta_{bc} \delta_{il} \delta_{jk} \\
\langle \tilde{\Phi}_{ij}^{ab} | \Psi^{(1)} \rangle &= T_{ab}^{ij} \\
\langle \Psi^{(0)} | \hat{H} | \tilde{\Phi}_{ij}^{ab} \rangle &= (ai|bj) = K_{ab}^{ij}
\end{aligned} \tag{2.31}$$

where K_{ab}^{ij} is an exchange integral. Taking into account the expression for the energy $E^{(2)}$ in equation (2.26), one can then define the MP2 energy contribution as

$$\begin{aligned}
E^{(2)} &= \langle \Psi^{HF} | \hat{H}^{(1)} | \Psi^{(1)} \rangle \\
&= \langle \Psi^{HF} | \hat{H} - \hat{H}^{(0)} | \Psi^{(1)} \rangle \\
&= \sum_{ijab} \langle \Psi^{HF} | \hat{H} | \tilde{\Phi}_{ij}^{ab} \rangle \tilde{T}_{ab}^{ij} \\
&= \sum_{ijab} K_{ab}^{ij} \tilde{T}_{ab}^{ij}.
\end{aligned} \tag{2.32}$$

The amplitudes are then calculated by solving the first order perturbation equations³¹

$$\begin{aligned}
R_{ij}^{ab} &= \langle \tilde{\Phi}_{ab}^{ij} | \hat{H}^{(0)} - E^{(0)} | \Psi^{(1)} \rangle + \langle \tilde{\Phi}_{ab}^{ij} | \hat{H} | \Psi^{(0)} \rangle = 0 \\
R_{ij}^{ab} &= K_{ab}^{ij} + \sum_c (f_{ac} T_{cd}^{ij} + T_{ac}^{ij} f_{cb}) - \sum_k (f_{ik} T_{ab}^{kj} + T_{ab}^{ik} f_{kj}) = 0,
\end{aligned} \tag{2.33}$$

where f_{mn} are elements of the Fock matrix. When canonical orbitals are used, $f_{mn} = \varepsilon_m \delta_{mn}$ and the previous expression is reduced to

$$T_{ab}^{ij} = -\frac{K_{ab}^{ij}}{(\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j)}. \tag{2.34}$$

Taking into account the second order energy present in equation (2.32), the canonical MP2 energy is then defined as

$$\Delta E^{MP2} = \sum_{ij} \sum_{ab} \frac{K_{ij}^{ab} (2K_{ij}^{ab} - K_{ji}^{ab})}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b}. \tag{2.35}$$

MP2 is known to retrieve about 80 to 90 % of the correlation energy for organic molecules.²⁷ MP2 presents a scaling of $\mathcal{O}(\mathcal{N}^5)$. This approach is one of the most used post-HF methods to date.^{26,27}

2.1.3 Local Correlation Methods

Dynamic correlation is a short range effect that decreases with r^{-6} as the dispersion energy. The high computational cost of treating this effect arises from the use of canonical orthogonal orbitals that are delocalized over the entire system. A way to overcome this effect is to take advantage of such short-range character and to treat correlation through local orbital spaces. For this effect Saebø and Pulay proposed the use of non-orthogonal virtual orbitals.^{17,18,32} It was demonstrated that with a moderate truncation of the configuration space, about 98 % of the correlation energy could be recovered when compared with the correspondent canonical approach.³³

Local molecular orbitals (LMO) can be obtained from the HF canonical orbitals by performing an unitary transformation as

$$\begin{aligned} |\phi_i^{loc}\rangle &= \sum_{\mu=1}^N |\chi_\mu\rangle L_{\mu i} \\ &= \sum_{k=1}^{M_{MO}} |\phi_k^{can}\rangle U_{ki}, \end{aligned} \tag{2.36}$$

where the canonical molecular orbitals are expressed as a linear combination of atomic orbitals. This transformation retains the requirement of orthonormality. Several localization schemes have been proposed, like the Pipek-Mezey³⁴(PM), Boys³⁵(B) and Edmiston-Ruedenberg³⁶(ER). In the B localization one maximizes the distance between orbital centroids, while in the ER method one maximizes the repulsion between orbitals.³⁵ Unsaturated bonds in the first normally appear as banana shaped, while the ER method gives conventional σ and π bonds. As in the case of the latter, the PM localization scheme is able to separate between σ and π orbitals by maximizing the orbitals partial charges. An alternative scheme using natural local atomic orbitals has been proposed.³⁷

The LMOs are projected from the AOs. The resulting orbitals are known as projected atomic orbitals (PAOs). The projection of the occupied space out of the AO space is given by³¹

$$\begin{aligned}
|\tilde{\chi}_r\rangle &= \left(1 - \sum_i^{M_{occ}} |\phi_i\rangle \langle \phi_i|\right) |\chi_r\rangle \\
&= \sum_{\mu} |\chi_r\rangle P_{\mu r},
\end{aligned}
\tag{2.37}$$

where \mathbf{P} is the projection matrix. The latter can be expressed as

$$\begin{aligned}
\mathbf{P} &= 1 - \mathbf{L}\mathbf{L}^\dagger\mathbf{S} \\
&= 1 - \mathbf{C}_{occ}\mathbf{C}_{occ}^\dagger\mathbf{S} \\
&= \mathbf{C}_{virt}\mathbf{C}_{virt}^\dagger\mathbf{S},
\end{aligned}
\tag{2.38}$$

where \mathbf{C}_{occ} and \mathbf{C}_{virt} are the MO coefficient matrices for the occupied and virtual space respectively and \mathbf{S} is the overlap matrix. This approach has the following properties. (1) All used orbitals are localized. (2) All the occupied orbitals are orthogonal between each other and with the virtual space. (3) The virtual orbitals are not orthogonal with an overlap matrix

$$\begin{aligned}
\langle \tilde{\chi}_r | \tilde{\chi}_s \rangle &= (\mathbf{P}^\dagger \mathbf{S} \mathbf{P})_{rs} \\
&= \tilde{\mathbf{S}}_{rs}.
\end{aligned}
\tag{2.39}$$

(4) Since the virtual space is not orthogonal, there must be linear dependencies.

Methods that take advantage of such localization procedure are known as local methods. In order to distinguish them from a canonical approach the prefix “L” is added. In this fashion the MP2 theory when used under a local approach is denominated as LMP2, as for CCSD will be LCCSD. This family of local methods is implemented in the Molpro quantum chemistry package.³⁸

In local methods one restricts the excitations from an occupied orbital to a nearby virtual orbital space. Such constraint will contribute to a lower scaling when compared with the respective canonical method. The subset of PAOs next to a localized molecular orbital i is called the domain of i ($[i]$). The selection of the PAOs can be achieved by several orbital domain selection criteria, like the method proposed by Boughton and Pulay (BP),³⁵ and the natural population analysis (NPA).³⁷ The latter exhibits a lower basis set dependency.³⁷

A single excitation from an occupied molecular orbital i to a virtual orbital s is only possible, if s belongs to the domain of i ($[i]$). In the same fashion, one can describe double excitations from occupied molecular orbitals i and j to virtual orbitals s and t ,

$ij \rightarrow st$, only if $st \in [ij]$, where $[ij] = [i] \cup [j]$. A consequence of such a truncation is that for a given pair the amplitudes do not increase quadratically with the system size, but instead become independent of the molecular size.³¹

Domain truncation schemes by itself do not impose linear scaling relative to the system size. Since electron correlation is a short-range effect, one would expect that orbital pairs that are very far apart have only a slight contribution to the correlation energy. This results in a second approximation in local methods. The orbital pairs are classified according to the distance between them and a hierarchical treatment is used.

The orbital pairs are divided in five classes: strong pairs, close pairs, weak pairs, distant pairs and very distant pairs. If one considers R_{ij} as the distance between two LMOs i and j , then $R_c > R_{ij}$ one has the strong pairs, when $R_w > R_{ij} \geq R_c$ one has close pairs. The weak pairs are considered in the interval $R_d > R_{ij} \geq R_w$ and distant pairs for $R_{vd} > R_{ij} \geq R_d$. Very distant pairs are those for which $R_{ij} > R_{vd}$. The constants R_c , R_w , R_d and R_{vd} are distances, with values of 1, 3, 8 and 15 Bohr respectively.³⁸ The very distant pairs are neglected since their contribution to the total correlation energy is negligible. The strong pairs are treated at the higher level of theory, i.e. in LMP2 these are treated at the MP2 level while in LCCSD this are treated at the CCSD level of theory. The remaining pairs are treated hierarchically with several approximations according to the local method. In the present work one made use only of the LMP2 method.

In order to obtain the LMP2 energy, one can define the first order wave function in the LMO and PAO basis as

$$|\Psi^{(1)}\rangle = \frac{1}{2} \sum_{ij \in P} \sum_{rs \in [ij]} T_{rs}^{ij} |\Phi_{ij}^{rs}\rangle$$

$$T_{rs}^{ij} = T_{sr}^{ji},$$
(2.40)

where P is the pair list which contains the pair domains $[ij]$. $|\Phi_{ij}^{rs}\rangle$ is a double excited determinant, while r and s are indices for the non-orthogonal PAOs (virtual space).³¹ Taking into account the residuals equation from canonical MP2 present in equation (2.33),

one can transform it into the PAO basis leading to³⁹

$$\begin{aligned}
R_{rs}^{ij} &= K_{rs}^{ij} + \sum_{tu \in [ij]} f_{rt} T_{tu}^{ij} \tilde{S}_{us} + \sum_{tu \in [ij]} \tilde{S}_{rt} T_{tu}^{ij} f_{us} \\
&\quad - \sum_k \left[\sum_{tu \in [kj]} \tilde{S}_{rt} f_{ik} T_{tu}^{kj} \tilde{S}_{us} + \sum_{tu \in [ki]} \tilde{S}_{rt} f_{kj} T_{tu}^{ik} \tilde{S}_{us} \right] \\
&= 0,
\end{aligned} \tag{2.41}$$

where the matrix \tilde{S}_{rs} is the overlap matrix in equation (2.39). In the local basis the Fock matrices f are not diagonal. In order to solve the doubles residual equation in the PAO basis for the amplitudes one needs to perform it iteratively.³¹ A more complete derivation from the LMP2 equations is available in Reference [31]. The second order energy can be written as

$$E^{LMP2} = \sum_{ij \in P} \sum_{rs \in [ij]} (2T_{rs}^{ij} - T_{sr}^{ij}) K_{ij}^{rs}, \tag{2.42}$$

where K_{ij}^{rs} are exchange integrals. The latter are given by

$$\begin{aligned}
K_{ij}^{rs} &= (ri|sj) \\
&= \int \phi_r(\mathbf{r}_1) \phi_i(\mathbf{r}_1) \frac{1}{r_{12}} \phi_s(\mathbf{r}_2) \phi_j(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2.
\end{aligned} \tag{2.43}$$

One should notice that in order to solve the residuals one only needs to solve a small subset of exchange integrals in the AO basis as⁴⁰

$$K_{ij}^{rs} = \sum_{\mu} P_{\mu r} \sum_{\lambda} P_{\lambda s} \sum_{\nu} L_{\nu i} \sum_{\sigma} L_{\sigma j} (\mu\nu|\lambda\sigma) \tag{2.44}$$

which represent the most expensive part from the LMP2 calculations. But due to domain pairs and distance approximations the method runs much faster than canonical MP2. LMP2 analytical gradients have been developed by Werner *et al.*^{41,42} LMP2 with density fitting approximations has been shown to improve the speed of the calculations relative to the basis set size with negligible errors added.⁴³

Treating the occupied and virtual space with local approximations allows one to decompose the interaction energy through an analysis of the excitations. A scheme illustrating

this analysis is given in Figure 2.1.

Considering the interaction between two monomers, the interaction energy can be decomposed into:⁴⁴

- Intramolecular correlation effects that result from the double excitation from LMOs and the pair domains belonging to the same monomer, Figure 2.1-a)
- Dispersive effects, that result from single excitations from the LMOs of each monomer to the domain that belongs to the respective monomer, Figure 2.1-b)
- Dispersion-exchange effects that result from cross single excitations from the LMOs of each monomer to the virtual space from the other monomer, Figure 2.1-c)
- Ionic effects that result from the excitation of electrons in the occupied LMOs from monomer A to the corresponding virtual space in A and electronic transfer from the monomer B to the virtual space of A, i.e. $i \rightarrow [i]$ and $j \rightarrow [i]$, where $i \in A$ and $j \in B$, Figure 2.1-d)

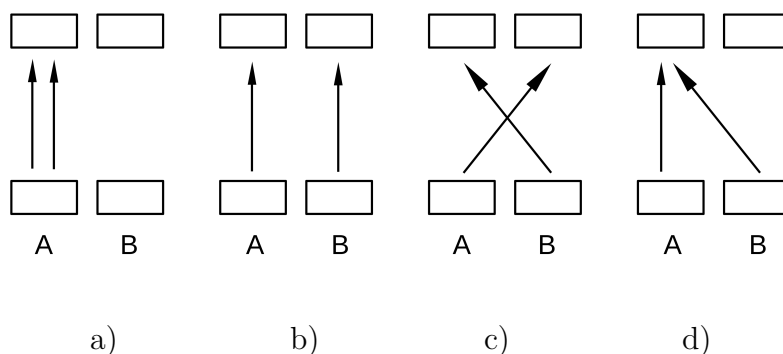


Figure 2.1: Schematic representation of possible double excitations in local correlation methods applied to intermolecular interactions between two monomers A and B. The rectangles represent ground and excited states of these two monomers and the arrows excitations. The several components in the interaction energy are: a) intramolecular correlation, b) dispersion, c) dispersion exchange and d) ionic effects

Double ionic substitutions from the type $j \rightarrow [i]$ and $j' \rightarrow [i]$ as well as ones from the type $i \rightarrow [i]$ and $i' \rightarrow [j]$ are excluded by construction. The schematics of such effects that are not taken into account in local methods are represented in Figure 2.2. The first is the main responsible for basis set superposition errors (BSSE).⁴⁴ Neglecting this

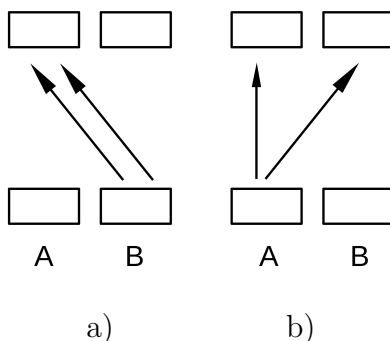


Figure 2.2: Schematic representation of possible double excitations that are excluded from local correlation methods applied to intermolecular interactions between two monomers A and B. The rectangles represent ground and excited states of these two monomers and the arrows excitations. The components in the interaction energy are: a) BSSE and b) ionic.

last two effects may result in a reduction of the correlation energy up to 10 % when comparing LMP2 with MP2.⁴⁴

2.1.4 Density Functional Theory

The methods presented so far are known as wavefunction methods, an alternative to the latter is density functional theory (DFT). The main motivation of the latter is to be able to calculate the energy of a system avoiding an explicit solution of the Schrödinger equation. In order to achieve this, DFT is focused in determining the energy of a system using a physical observable.

In wavefunction methods the Hamiltonian depends on the positions of the nuclei and the total number of electrons. The dependence of the latter seems to indicate that the use of the electron density, $\rho(\mathbf{r})$, is a plausible observable, since the integration over all space gives the total number of electrons of the system, N ,

$$N = \int \rho(\mathbf{r}) d\mathbf{r}. \quad (2.45)$$

Through the knowledge of the density one can also define the nuclei, since in their position ρ presents a local maxima. In DFT the electrons interact between themselves and with an external potential, V_{ext} .

In order to define the energy in terms of the density, one needs to take into consideration the Hohenberg-Kohn theorems.²⁶ The first also known as the Hohenberg-Kohn existence theorem,²⁶ implies that the non-degenerated ground-state density determines

not only the external potential but also the Hamiltonian and therefore the wavefunction. The second is a variational theorem, that states that any trial density gives an upper bound to the exact ground-state energy. Only the exact ground-state density will give the exact ground-state energy. The main inconvenience in this approach is the rational choice of the density.

Further approximations were proposed by Kohn and Sham. They suggested the use of an operator for a non-interacting system, defined as a sum of one electron operators. The key point in their approximation is to take as starting point a fictitious system of non-interacting electrons with a ground state density that is the same as in the real system. The energy functional is then defined as

$$E[\rho(\mathbf{r})] = T_{ni}[\rho(\mathbf{r})] + V_{ne}[\rho(\mathbf{r})] + V_{ee}[\rho(\mathbf{r})] + \Delta T[\rho(\mathbf{r})] + \Delta V_{ee}[\rho(\mathbf{r})], \quad (2.46)$$

where T_{ni} is the kinetic energy of non-interacting electron, the following two terms are the classical nuclear-electron and electron-electron interactions and the two last terms are a correction to the kinetic energy between the fictitious and real system. The sum of the last two terms is the exchange-correlation energy, E_{xc} . Making use of a Slater determinant, the density can be defined as²⁷

$$\rho = \sum_i^N \langle \phi_i | \phi_i \rangle. \quad (2.47)$$

One can rewrite equation (2.46) in an orbital formalism as

$$\begin{aligned} E[\rho(\mathbf{r})] = & -\frac{1}{2} \sum_i^N \langle \phi_i | \nabla_i^2 | \phi_i \rangle - \sum_i^N \sum_k^M \langle \phi_i | \frac{Z_k}{|\mathbf{r}_i - \mathbf{r}_k|} | \phi_i \rangle \\ & + \frac{1}{2} \sum_i^N \langle \phi_i | \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' | \phi_i \rangle + E_{xc}[\rho(\mathbf{r})], \end{aligned} \quad (2.48)$$

where N is the number of electrons, M the number of nuclei, and $\rho(\mathbf{r}')$ is the density for the non-interacting system. As in HF the next step is to find the orbitals that minimize the energy. This leads to the pseudo-eigenvalue equations

$$h_i^{KS} |\phi_i\rangle = \varepsilon_i |\phi_i\rangle, \quad (2.49)$$

which need to be solved iteratively, where h_i^{KS} is the Kohn-Sham one-electron operator defined as

$$h_i^{KS} = -\frac{1}{2}\nabla_i^2 + \sum_k^M \frac{Z_k}{|\mathbf{r}_i - \mathbf{r}_k|} + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \frac{\delta E_{xc}}{\delta \rho}. \quad (2.50)$$

Since the density of the non-interacting system is identical to the one for the real system, the kinetic energy of the fictitious system will be very close to the real system.

In the limit of the exact E_{xc} , DFT is an exact theory. Unfortunately, the exact form of the exchange functional is not known. Over the years several formulations of the exchange functional were proposed. In some of these, the value of E_{xc} in a given position is computed exclusively from the value of the density in that position, this is known as local density approximation (LDA).^{26,45} In this family of functionals the exchange and correlation kernel are derived from the uniform electron gas.²⁷ The correlation functional was not deduced from first principles but instead, obtained from Monte Carlo simulations. Between the most used LDA correlation functionals are the ones from Vosko *et al.*⁴⁶ and Perdew.⁴⁷

In a molecular system the density is far from being uniform and it is expected that LDA faces some limitations. In order to correct the locality of the previous functional, a term has to be introduced which will take into account how the density is locally changing. The latter information is given by the gradient of the density. This is known as generalized gradient approximation (GGA).^{26,45} Some examples from the latter include the B88,⁴⁸ and PBE⁴⁹ functionals for exchange and the PBE⁴⁹ and LYP⁵⁰ for the correlation.

The exchange and correlation functionals can be combined giving rise to a high variety of methods. Between these one of the most famous is the B3LYP.⁵¹ The latter has been applied to a huge variety of systems and is so successful that some authors even tend to mention it only as DFT.⁴⁵ The exchange correlation functional is defined as follows

$$E_{xc}^{B3LYP} = 0.2E_x^{HF} + 0.72E_x^{B88} + 0.08E_x^S + 0.81E_c^{LYP} + 0.19E_c^{VWN80}, \quad (2.51)$$

where the exchange is composed by the exact HF exchange, E_x^{HF} , and by the B88 and Slater-Dirac.⁵² The correlation part is composed by the LYP and VWN80⁴⁶ functionals. This functional is known as a hybrid functional, since it makes use of Hartree-Fock exchange.

Although DFT can be applied to a large variety of systems, the current functionals are

known to have some disadvantages. They are unable by construction, to describe weak interactions like dispersion forces. In order to overcome this, empirical based corrections to the DFT functionals have been proposed with significant success.^{53–55} DFT is also known to poorly describe charge transfers, due to the local character of the exchange-correlation functional.²⁷ Contrary to wavefunction methods, there is not a well defined hierarchy of methods which converge to the exact total energy of the system.

2.2 Molecular Mechanics

The methods presented so far have been approximations to solve the Schrödinger equation. Under the Born-Oppenheimer approximation, one is interested in the electronic structure of the system. However, due to the high computational cost involved, these can not be applied to systems with thousands of atoms. In this way quantum mechanical calculations are not applicable to biomolecular systems, such as proteins.

In order to be able to compute systems with such large dimensions one can use molecular mechanics (MM) instead of quantum mechanics. In the former the energy is described by classical physics, parametrizing the degrees of freedom in the system. In other words the total energy is given by the sum of the energies resulting from bond stretching, angles, torsion and non-bonded contributions. This expression is parametrized according to experimental and to QM calculations performed in smaller systems. The resulting energy expression is referred as a force field (FF).

There are a high variety of FFs.²⁶ They are normally separated in two major categories, the all-atom and united-atom force fields. The former treat explicitly all the atoms of the system in study, while in the latter there is a formation of a pseudo-atom which contains the information of non-polar hydrogens and the heteroatom to which they are bonded. In other words, in the united atom FFs some atoms are grouped into one that describes the properties of all the set. As an example, the optimized potentials for liquid simulations (OPLS) force field is presented. The latter is parametrized for biomolecules and some organic compounds.^{56–58}

$$E_{OPLS} = E_{bond} + E_{angles} + E_{torsion} + E_{non-bonded} \quad (2.52)$$

The total energy is defined as a sum of several terms that describe properties of the system. These are the bond stretches (E_{bond}), bond angles (E_{angles}), torsions ($E_{torsion}$)

as well as non-bonded terms ($E_{non-bonded}$). The latter include van der Waals and electrostatic interactions. In the case of the OPLS these terms are represented as follows:

$$\begin{aligned}
 E_{bond} &= \sum_{bonds} K_r (r - r_{eq})^2 \\
 E_{angles} &= \sum_{angles} K_\theta (\theta - \theta_{eq})^2 \\
 E_{torsion} &= \sum_i \frac{V_1^i}{2} [1 + \cos(\phi_i + f_{i1})] + \frac{V_2^i}{2} [1 + \cos(2\phi_i + f_{i2})] + \frac{V_3^i}{2} [1 + \cos(3\phi_i + f_{i3})] \\
 E_{non-bonded} &= \sum_i^A \sum_j^B \left[\frac{q_i q_j e^2}{r_{ij}} + 4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^6}{r_{ij}^6} \right) \right] f_{ij}.
 \end{aligned} \tag{2.53}$$

The first two terms are parametrized through an harmonic potential. In both a force constant, K , needs to be fitted as well as an equilibrium distance, r_{eq} or angle θ_{eq} , which depend on the atoms type. Under this formalism bond breaking is not taken into account. The torsion potential describes the contribution of the dihedral angles, ϕ_i . In the former contribution, V are coefficients in the Fourier series, while f_{i1} , f_{i2} and f_{i3} are phase angles. The non-bonded contribution is defined by Coulomb and van der Waals terms between molecules A and B , with standard combination rules,

$$\begin{aligned}
 \sigma_{ij} &= \sqrt{\sigma_{ii}\sigma_{jj}} \\
 \epsilon_{ij} &= \sqrt{\epsilon_{ii}\epsilon_{jj}}.
 \end{aligned} \tag{2.54}$$

The van der Waals interactions are represented by a 12-6 Lennard-Jones potential. The non-bonded interactions occur between all pair of atoms ($i < j$) separated by three or more bonds. f_{ij} parameter is equal to one but for 1-4 intramolecular interactions it presents the value of 0.5.

The use of molecular mechanics is quite useful for example in conformational studies. In order to retrieve a reliable information of the system one needs many configurations. The latter can be obtained through molecular dynamics or Monte Carlo simulations.

Molecular dynamics simulations (without the use of thermostats or barostats) will run under a microcanonical ensemble (NVE), this is at a fixed number of particles (N), volume (V) and total energy (E). The latter is composed by the sum of the kinetic and potential energy. In MD the system evolves along a time coordinate and new configurations are obtained by solving Newton's equations for motion under the law of

conservation of energy.

In Monte Carlo simulations are carried out mostly with a canonical ensemble (NVT), where T stands for temperature. In the former unlike molecular dynamics, the configurations of the system are generated by random trial movements. Since one does not solve Newton's equations for motion, this permits the decoupling of the kinetic energy from the total energy of the system. The kinetic term can then be added a posteriori. Under this ensemble the simulation is carried out at constant temperature, and unlike molecular dynamics there is no need for additional thermostat algorithms.⁵⁹

One of the main drawbacks of Monte Carlo comes from the fact that one does not follow the evolution of the system over time. In this way one lacks the information about the dynamics of the system. As a result Monte Carlo focuses in retrieving thermodynamic and physical properties of a system in equilibrium.

Monte Carlo algorithms are also known to be embarrassingly parallel, this means that they can easily run in parallel in CPUs clusters or even GPUs.

The results obtained by any MM simulation method are analysed through the use of statistical mechanics. Since in a MM approach a great number of configurations are generated, enough disk space to keep this information can become an issue. Nonetheless, this approach is quite successful in the study of big biologic systems in solution.

In the next section one will present a general description of the Metropolis Monte Carlo method.

2.2.1 Metropolis Monte Carlo

In the Metropolis Monte Carlo (MC)⁶⁰ method constrained to a canonical ensemble, a certain average property A ($\langle A \rangle$) can be calculated, according to classical statistical mechanics, as⁶¹

$$\langle A \rangle = \frac{\int A(\mathbf{r}^N) e^{-\beta U(\mathbf{r}^N)} d\mathbf{r}^N}{\int e^{-\beta U(\mathbf{r}^N)} d\mathbf{r}^N} = \frac{\sum_{i=1}^{M_{tot}} A(\mathbf{r}_i^N) e^{-\beta U(\mathbf{r}_i^N)} d\mathbf{r}_i^N}{\sum_{i=1}^{M_{tot}} e^{-\beta U(\mathbf{r}_i^N)}}, \quad (2.55)$$

with

$$\beta = \frac{1}{k_B T}, \quad (2.56)$$

where U is the potential energy, \mathbf{r}_i^N is the configuration and k_B the Boltzmann constant. The denominator of equation (2.55) is commonly designated as the configurational integral, Z . The main problem for the calculation of the average property is the calculation

of the latter integral.

An elegant way to solve the problem is to consider that the probability density of finding the system in a certain configuration \mathbf{r}^N is defined by the Boltzmann distribution. In this way the previous mentioned probability can be given as

$$\rho(\mathbf{r}^N) = \frac{e^{-\beta U(\mathbf{r}^N)}}{Z}. \quad (2.57)$$

Let us consider a integral of a property F composed by

$$F = \int_a^b f(\mathbf{x}) d\mathbf{x} = \int_a^b \frac{f(\mathbf{x})}{\rho(\mathbf{x})} \rho(\mathbf{x}) d\mathbf{x}, \quad (2.58)$$

where $\rho(\mathbf{x})$ is an arbitrary probability density. If one needs to solve it numerically by performing trial attempts, τ_{max} where each trial τ is an aleatory number between a and b and with a distribution $\rho(\tau)$, then we can rewrite the previous integral as

$$F \simeq \frac{1}{\tau_{max}} \sum_{\tau=1}^{\tau_{max}} \frac{f(\tau)}{\rho(\tau)}, \quad (2.59)$$

where $f(\tau)$ is the property f in the trial τ . Taking into consideration equations (2.55) and (2.59) and the fact that the density of probability of a certain state presents the form of a Boltzmann distribution, one can see that this avoids the need to calculate the integral Z . In this way in MC the property A can be defined approximately as an average value of A over all generated configurations as

$$\langle A \rangle \approx \frac{1}{M_{tot}} \sum_{i=1}^{M_{tot}} A(\mathbf{r}_i^N), \quad (2.60)$$

where M_{tot} is the total number of MC configurations and $A(\mathbf{r}_i^N)$ is the property measured in a certain configuration \mathbf{r}_i^N . In this approach one does not know the absolute probability of visiting a certain configuration but only the relative probability.

In MC one starts by defining a configuration, which will be referred to as old (o). In the next step a new configuration (n) is generated by performing a random displacement. The new configuration is then accepted or neglected according to a certain criterion.⁶¹

If one considers a transition from an old state to a new state, that preserves the system in equilibrium, the former will be associated with a transition probability of π_{on} defined

as

$$\pi_{on} = \alpha_{on}\gamma_{on}. \quad (2.61)$$

The π_{on} depends on the probability of accepting the trial configuration (γ_{on}) and from a transition matrix α_{on} which is related with the so called Markov chain of states. The latter in the Metropolis method is symmetric, where $\alpha_{on} = \alpha_{no}$.

The Markov model is a probabilistic model that is applied to systems that present a specific type of dependency, this is, the state of the system in the configuration i presents only a dependency from the previous state, $i - 1$. In this way a future configuration depends only on the current state and not how the current state was obtained. In this way MC is a memoryless procedure.

For a system in equilibrium one can write that the ratio between the probability of the new state and the old state is given by⁶¹

$$\frac{\rho_n}{\rho_o} = \frac{\gamma_{on}}{\gamma_{no}} = e^{-\beta[U(n)-U(o)]}. \quad (2.62)$$

If the new state presents a lower energy then the previous state then the new configuration is accepted. If this is not true then one proceeds as follows. According to equation (2.62) the new configuration should be accepted with a probability of

$$\gamma_{on} = e^{-\beta[U(n)-U(o)]} < 1. \quad (2.63)$$

The next step involves the generation of a random number within a uniform distribution in the interval $[0, 1]$. If the random number is lower than γ_{on} in equation (2.63) then the new configuration is accepted. If the new state does not obey the two previously mentioned restrictions then it is rejected.

So far nothing has been said in relation to the random move in a Monte Carlo step. If one assumes that the molecules present a fixed geometry, one can define the trial movements in terms of translation and rotation, for example in relation to the center of mass.

In order to create a translational movement, one starts by creating a random vector (e_x, e_y, e_z) where $e_x^2 + e_y^2 + e_z^2 = 1$. Then one creates two random numbers ξ_1 and ξ_2 independent and uniform on an interval $[-1, 1]$ until $S = \xi_1^2 + \xi_2^2 < 1$. The unitary vector

is then defined as⁶²

$$(e_x, e_y, e_z) = (2\xi_1\sqrt{1-S}, 2\xi_2\sqrt{1-S}, 1-2S). \quad (2.64)$$

The new coordinates can then be defined as⁶²

$$\begin{aligned} x^{new} &= x^{old}(\delta_{max}e_x) \\ y^{new} &= y^{old}(\delta_{max}e_y) \\ z^{new} &= z^{old}(\delta_{max}e_z), \end{aligned} \quad (2.65)$$

where δ_{max} is a random number in an uniform distribution between zero and the maximum displacement allowed. If the latter is too small, it will lead to an increase of the number of accepted configurations but to a very poor sampling from the phase space. This is, consecutive states would be highly correlated and one would need a higher number of MC steps to explore the phase space. If the maximum displacement is too large, then most of the trial moves are rejected and the accepted configurations do not really represent the phase space.⁶³ Some authors mention that usually the maximum displacement should be defined in such a way so as to obtain 50 % of accepted steps, although this percentage is not clear to be the optimum choice.^{61,63}

A MC procedure needs to be ergodic, this means that every state in the phase space should be reached by a finite number of steps from any other state.⁶¹ Monte Carlo algorithms have also been carried out with the other ensembles like the isothermal-isobaric (NPT), grand-canonical (μVT). In the previous ensembles P stands for pressure and μ for the chemical potential.⁶³

The NPT is mostly applied when the system in study is expected to suffer a change of state. This is, to create two phases with different density. In this case the use of the canonical ensemble would not be a rational choice, since the system under this ensemble is confined to a fixed density.⁶¹

The use of the grand-canonical ensemble is the most reasonable choice when one is interested in understand the average behaviour of particles under the influence of an external property. As for example to study adsorption phenomena in heterogeneous catalysis as a function of the temperature.

2.3 Quantum Mechanics/Molecular Mechanics

The main drawback in MM approaches is related to their inability to properly describe chemical reactions, where bond breaking and bond formation are involved. The same is not true for QM methods. Normally the region in space where significant changes in the electronic structure occur is quite small. As example of such cases are enzymes. In the latter one uses the concept of an active center, with the rest of the enzyme keeping its structural properties. For these systems it would be convenient to be able to treat the active site with accurate electronic structure methods, while for the rest of the super-system a model with lower level of complexity could be used. In this way one would divide the system in two parts, one which is treated at higher level by QM and another at the MM level. The interface between the two regions is defined as an hybrid of QM and MM.⁶⁴ This formalism is known as QM/MM approach. The total Hamiltonian of the system can be translated in a simple fashion as

$$\hat{H}_{tot} = \hat{H}_{QM} + \hat{H}_{MM} + \hat{H}_{QM/MM}, \quad (2.66)$$

where \hat{H}_{QM} accounts for all interactions of the quantum mechanical particles with each other at the QM level. \hat{H}_{MM} accounts for the classical interaction between MM particles and $\hat{H}_{QM/MM}$ accounts for all the interactions between the quantum mechanical particle and the classical system.²⁶

Several formalisms have been proposed to carry out QM/MM calculations. In the case where the QM is polarized and the MM part is unpolarized the $\hat{H}_{QM/MM}$ can be described according to

$$\hat{H}_{QM/MM} = - \sum_i^{N_{se}} \sum_m^M \frac{q_m}{r_{im}} + \sum_k^{N_{sn}} \sum_m^M \left[\frac{Z_k q_m}{r_{km}} + 4\epsilon_{km} \left(\frac{\sigma_{km}^{12}}{r_{km}^{12}} - \frac{\sigma_{km}^6}{r_{km}^6} \right) \right], \quad (2.67)$$

where the QM part is composed by N_{se} electrons and N_{sn} nuclei and the MM region is composed of M atoms. In the previous equation q_m stands for the charge of a MM atom, r_{im} is the distance between a QM electron and a MM atom and r_{km} is the distance between an atom QM and a MM atom. In this case the van der Waals interactions between the two regions is described as a 12-6 Lennard-Jones potential. The first two terms are the electrostatic interaction between the two regions.²⁶

Under this formalism the QM is polarized due to the MM charges (also known as

electrostatic embedding), but the reverse is neglected. This approach can be considered unbalanced. This can be overcome through the use of polarizable force fields. The latter are more elaborated in terms of implementation.

Let us consider large systems, where one defines two regions, one small and one large, in which the latter is to be treated at the MM level, while the small more interesting part of the system will be treated at the QM level. In this approach the energy of the whole system can be expressed as a linear combination of model compounds with different sizes and levels of theory, defined as

$$E_{total} = E_{MM}^{large} + (E_{QM}^{small} - E_{MM}^{small}) \quad (2.68)$$

where the large system is treated only at the MM level and the small region, also known as core, is treated at the QM and MM level. The third term in equation (2.68) is to avoid double counting. This formalism is known as subtracting scheme. Since the calculations are decoupled this allows one to have more than two regions. Within this approach it is possible, for example, to have the QM region separated in subregions each one treated with different levels of theory.¹³ This methodology is known as ONIOM.⁶⁵ In its initial formalism E_{QM}^{small} was obtained in vacuum and the polarization of the small part by the rest of the molecule was neglected. The latter effect was overtaken through the inclusion of electrostatic embedding.⁶⁶ Another example of a subtracting scheme is the so called electrostatic embedded generalized molecular fractionation with conjugates caps (EE-GMFCC). In the latter the system is fractioned in smaller parts and one makes use of capping fragments.⁶⁷ The energy of the system is then calculated as the difference of the energy from the sum of the individual capped parts and the sum of the caps.

2.4 Implicit Solvation Models

In this thesis one proposes to study phenomena that occur in solution. In computational chemistry the effects of the latter can be simulated through the use of explicitly or implicitly solvation models. In the first case one has a full atomistic description, while in the latter the influence of the environment is included with a continuum medium. Although the latter can be applied to describe any kind of media, in the following manuscript one is more interested in its applicability to describe a solvent. Taking this into account, one will provide a general overview of implicit solvation models. We will

focus in its application to quantum mechanics.

Under the Born-Oppenheimer approximation one can define the effective Hamiltonian as $\hat{H}_{eff} = \hat{H}_0 + \hat{H}_{env}$. Hereby \hat{H}_0 is the Hamiltonian of the solute in gas phase and \hat{H}_{env} describes the environment. In the case of continuum solvation models (CSMs), $\hat{H}_{env} = V_{cont}$, which depends only on the coordinates of the solute.^{68,69} When the effective Hamiltonian under the previous formalism is used in HF or DFT, it can be solved as a regular self-consistent reaction field (SCRF) problem.^{68,69}

Upon solvation the wavefunction of the solute interacts with the charge distribution from the solvent. In a continuum approach the latter is substituted by apparent charges. The latter result as a response of the dielectric continuum to the presence of the solute, thereby including a statistical average of the solvent degrees of freedom in equilibrium. One could say that under this approximation the solvent exists because the solute exists. In CSMs the solute and solvent have a mutual polarization effect.⁶⁸

The foundations of the continuum solvent models is the description of the electrical potential by a Poisson equation of classical electrostatics. The latter is defined as a function of the electron density, $\rho(\mathbf{r})$, and the dielectric constant, ε , of the medium as

$$\nabla^2 V_{ele}(\mathbf{r}) = -\frac{4\pi\rho(\mathbf{r})}{\varepsilon}. \quad (2.69)$$

In a CSM the solute is thought to be inside of a cavity that is defined by the ε . Taking into account the two regions, equation (2.69) can be rewritten as

$$\nabla\varepsilon(\mathbf{r}) \cdot \nabla V_{ele}(\mathbf{r}) = -4\pi\rho(\mathbf{r}). \quad (2.70)$$

The previous equation is valid when the solution ionic strength is zero. When electrolytes are taken in consideration, then one makes use of the Poisson-Boltzmann (PB) equation

$$\nabla\varepsilon(\mathbf{r}) \cdot \nabla V_{ele}(\mathbf{r}) - \varepsilon(\mathbf{r})\lambda(\mathbf{r})\kappa^2 V_{ele}(\mathbf{r}) = -4\pi\rho(\mathbf{r}), \quad (2.71)$$

where λ is zero or one, depending if the region is accessible to the electrolyte or not, and κ^2 is the Debye-Hückel parameter defined as $8\pi q^2 I / \varepsilon k_B T$. q is the charge of the electrolyte, I the ionic strength, T the temperature and k_B the Boltzmann constant. For ideal cavities the Poisson-Boltzmann equation has a simple resolution.

There are a high variety of CSMs. The latter can be separated in six categories ac-

cording to their electrostatics: the apparent surface charge methods; multipole expansion methods; generalized Born approximation; image charge methods; the finite elements methods and the finite element methods.⁵ Belonging to the first group are the polarizable continuum model (PCM)^{5,8,9,70-72} and the conductor-like screen model (COSMO).^{5,6} The latter is the implicit solvation model used in the present work.

Under such approaches the cavity of the molecular system can be considered to be formed by overlapping spheres from the constituting atoms, considering the van der Waals radii of the latter. The cavities formed are not considered ideal and the PB equation needs to be solved numerically. The determination of the cavity involves a construction of a grid followed by the construction of the molecular cavity. Then one discretizes the solute charge distribution into the inside of the grid points. Lastly, one calculates the electrostatic potential at each point of the grid and the charge-potential integral²⁶

$$V = -\frac{1}{2} \int \rho(\mathbf{r}) V_{ele}(\mathbf{r}) d\mathbf{r}. \quad (2.72)$$

In order to facilitate calculations the surface is tessellated into triangles, and the charge density collapsed in the center.⁷³ Some instability can occur from short charge-charge interactions. To avoid this some regions can be deleted near the overlap of the spheres.²⁶ In Figure 2.3 one presents an example of a solute in a cavity surrounded by charges that result as a response of the medium to the presence of the solute.

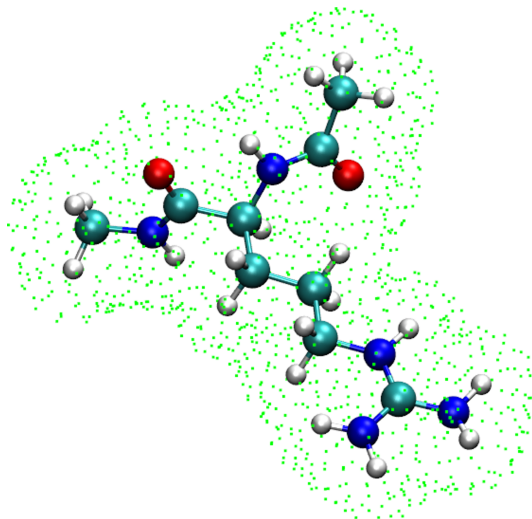


Figure 2.3: Protonated arginine involved in a cavity with COSMO charges on the surface.

The target electronic structure method is then solved in the presence of the potential. In the case of the Fock equations, this would take the form²⁶

$$(f_i - V) \psi_i = e_i \psi_i, \quad (2.73)$$

with the potential being an effective one-electron operator.

In COSMO the charges on the tessellated molecular surface are determined using a Poisson equation using a volume element approach.^{6,26}

Although PCM and COSMO seem to have different approaches both give very similar results. The COSMO approach was thought only to give appropriate results for solvents with high dielectric constants, but several studies have found a good agreement also in the case of apolar solvents.⁷⁴

The electrostatic component of the solvation free energy is not physically observed.²⁶ So it is impossible to discuss if one CSM algorithm is more or less accurate in terms of this component. The non-electrostatic component, as the previous, can not be measured. In this way some experimental results can be biased when one effect or another dominates. The non-electrostatics in CSM is assumed to be a characteristic of the atoms and is proportional to their solvent accessible surface area (SASA).²⁶ The treatment of dispersion in CSM is still a matter of investigation. It has been proposed to use a single scaled approach based on high semi-classical parametrization in conjunction with the PCM model, which revealed quite promising.⁷⁵

CSMs can be considered faster than explicit solvation models, since they include by construction a configurational sampling of the solvent avoiding the need of a statistical treatment of many solvent configurations, as it happens in the latter.⁷⁶ In the CSM the SCRF part adds normally about 15 % to the total time of an SCF calculation. Nonetheless, the speed up in the computational time is not an advantage if the result reveals to be incorrect. CSMs due to the lack of an atomistic description of the solvent, perform poorly when specific interactions between the solute and solvent are involved, like hydrogen bonds. Nonetheless, in general a good agreement between the two approaches have been shown to exist.^{68,73,77} CSMs have been successfully applied to the study of activation energies in S_N2 reactions of acetate ions and ethyl halide with aprotic and protic solvents.⁷⁸ The authors report a maximum deviation of 2.6 kcal·mol⁻¹ from experimental results.⁷⁸ Pliego *et al.* demonstrated that the extra addition of some explicit solvent molecules to the solute improves the entropy and free energy of solvation.⁷⁹ The

previous author also applied the same methodology to assess the pK_a of several organic and inorganic acids, where they obtained a better agreement with experimental results than in the case where the explicit solvent molecules were absent.⁸⁰ The inclusion of the latter seems to improve the performance of the CSMs. Nonetheless one question still remains. How many explicit solvent molecules are necessary to include according to a specific system? One possible way to choose the position of the solvent molecules around the solute would be to make use of molecular mechanics. Nonetheless one would need to choose several configurations in order to obtain a reasonable picture of the solvent influence.

3 Local MP2 in combination with COSMO

In the first part of this thesis one discusses the implementation of COSMO-LMP2 analytical gradients in the Molpro software package.³⁸ The performance of the method was then evaluated towards geometry optimizations of several benchmark systems in solution. In the latter set one also considers the interaction energy between a charged amino acid model and halide anions and water. One also assessed the behaviour of such approach towards conformational studies of a larger system. The latter was composed of a β -peptide tetramer with 74 atoms.

3.1 Method

One will start by discussing the COSMO model. The most marking difference between COSMO and other CSMs, is in the approximation condition to describe the dielectric continuum. The former uses a scaled conductor to describe the effect of the continuum, under a vanishing boundary condition. The solute is then contained inside of a cavity that results from interlocking spheres. The latter are formed taking into account the van der Waals radii of the atoms of the solute. The surface of the cavity is then charged, as the result from the induced polarization due to the presence of the solute. The charges on the former are denominated apparent screen charges. The surface is then tessellated into several segments in order to facilitate the computation of the electrostatic potential. If one represents the latter with a vector \mathbf{V}^X , for a certain solute X , the boundary condition can be represented as

$$0 = \mathbf{V} = \mathbf{V}^X + \mathbf{A}\mathbf{q}^*, \quad (3.1)$$

where \mathbf{A} is the Coulomb interaction matrix of the screening charges, and \mathbf{q}^* the screening charges of a true conductor. The latter is verified for $\epsilon = \infty$. In order to describe a finite dielectric, as in the case of usual solvents, there is the need to scale the screening charges. The expression that traduces the relation of the solute electrostatic potential and the screening charges \mathbf{q} on the surface of the cavity is given by

$$\mathbf{A}\mathbf{q} = -\frac{\epsilon - 1}{\epsilon + x}\mathbf{V}\mathbf{x} = -f(\epsilon)\mathbf{V}\mathbf{x}. \quad (3.2)$$

In the previous equation $f(\epsilon)$ is a scaling factor, which is dependent on the permittivity of the solvent. Normally x is set to 0.5 by default.

The solute electrostatic potential is defined by the charge density that is described in a QM calculation. At the HF level the solvent polarization can be carried out by adding to the gas phase Fock operator $\mathbf{F}^{g,n}$ at each SCF cycle n , the external potential \mathbf{Q} of the screening charges \mathbf{q} . \mathbf{Q} is determined by the charge density $\mathbf{P}^{(n-1)}$ of the $(n - 1)$ th cycle. This can be represented as

$$\mathbf{F}^{(s,n)} = \mathbf{F}^{(g,n)} + \mathbf{Q}(\mathbf{P}^{(n-1)}). \quad (3.3)$$

In the previous equation the superscript s stands for solution, while g stands for gas phase. $\mathbf{F}^{(s,n)}$ is the Fock operator used to calculate the Hartree-Fock solution. The use of CSMs at the HF level is well established. Nonetheless the inclusion of the former at a perturbative level is still doubtful. There are three possible ways to tackle the problem.^{70,81–88}

The first is usually mention as the non-iterative energy only scheme (PTE). Under such approach one starts to converge the Fock operator from a COSMO-HF calculation. The electron correlation treatment is then performed with the Fock operator and canonical orbitals from the latter mentioned calculation. In this way the screening charges are exclusively defined at the HF level.

A second alternative would be to recalculate the screening charges on the basis of the MP2 density, generating a new effective potential. The latter would be then reintroduced in a new COSMO-HF calculation. This procedure is then repeated until energy convergence within a threshold. This methodology is denominated as iterative energy and density scheme (PTED).

The third alternative can be seen as a truncation of the previous scheme, known as PTD scheme. Under this approach the new potential generated by the MP2 density is kept frozen and reused in a HF and MP2 calculation.

Ángyán *et al.*, pointed out that the only scheme which provides an energy correction consistent with the Wigner’s $(2n + 1)$ rule is the simplest PTE scheme.^{83,89} One will therefore give focus on the PTE model, and the remaining of this work will reference exclusively this approach, the latter being mentioned as COSMO-LMP2 throughout this manuscript.

In the PTE framework, the treatment of electron correlation is decoupled from the COSMO perturbation. In this way the extra gradient terms are derived in the same fashion as in the PCM gradient. The latter is present in Reference [70]. The main difference lies in the definition of the boundary condition. In this way one only needs to replace the PCM solvent operator for the COSMO equivalent (in equation (37) in the before mentioned reference). The variation in the Coulomb interaction $J_{R_\alpha}^{(2)}$ between the MP2 density ($\mathbf{P}^{(2)}$) and the converged COSMO-HF solution in respect to the nuclear movement R_α is then given by

$$J_{R_\alpha}^{(2)} = f(\epsilon) \left[\frac{1}{2} \mathbf{Q}(\mathbf{P}^{\text{HF}}) + \mathbf{Q}(\mathbf{P}^{(2)}) \right] \nabla_{R_\alpha} \mathbf{A} \mathbf{Q}(\mathbf{P}^{\text{HF}}). \quad (3.4)$$

The COSMO-LMP2 gradient is then computed in the same way as equation (3.4). The second order correction to the density matrix is computed by solving a set of coupled-perturbed Hartree-Fock equations (CPHF). For LMP2 one still needs to compute an additional set of coupled-perturbed localization (CPL) equations, taking into account the effect of the nuclear displacements on the LMO coefficients.⁴¹ The calculation of these coupled-perturbed equations are necessary for the LMP2 gradients in gas phase. To take into account the solvent effect one needs to perform the update of the boundary potential during the CPHF cycles, while the CPL equations are kept unchanged.

In the following section one will make use of such an approach in the study of several molecular systems.

3.2 Test systems

In order to test the viability of the previously presented method a series of benchmark calculations has been carried out for several molecules. The set was composed of water, ammonia, furan, ethanol, diethyl ether, cytosine, acetyl alcohol, glycine, water dimer, ammonia dimer and formic acid dimer.

Two solvents were considered, namely toluene and water. These were chosen since they present two extremes in terms of polarity. Toluene is an apolar solvent with a dielectric constant (ϵ) of 1.89 and water a polar solvent with a dielectric constant of 80.4. One should take into account that toluene presents a dielectric constant very close to some of the values usually applied when approximating the effect of an enzyme pocket (with $\epsilon = 3$ -5). Another set of test systems consisted of arginine interacting with chloride, bromide and water and another composed of a tetrapeptide of *cis*-2-amino-cyclopentane carboxylic acid (*cis*-ACPC) in two conformations, β -sheet and a 14-helix.

3.2.1 Computational Details

For the benchmark systems, structure optimizations were carried out with the cc-pVTZ Dunning basis sets⁹⁰ at the MP2 and LMP2 levels of theory in solution. For the latter two solvent were simulated, toluene and water, through the use of the COSMO solvation model.⁶ In the arginine interacting with the halide anions and water, the amino acid was acetylated in the N-terminus and methylaminated at the C-terminus, thereby removing the acidic and basic moieties of the backbone. This allows one to focus on the interactions with the guanidinium groups avoiding the interference of the terminal groups. The interacting species were placed in three possible docking sites in the plane of the guanidinium group as specified in Figure 3.1. The interacting systems were then optimized with the cc-pVDZ Dunning basis set⁹⁰ at the MP2 and LMP2 levels of theory. The optimizations were performed for gas phase and aqueous medium. The effects of the latter were introduced with the COSMO model with a $\epsilon = 80.4$.

As a peptide model one built a *cis*-ACPC tetrapeptide composed of 74 atoms in two conformations, β -sheet like and a 14-helix. The geometries were optimized with the cc-pVDZ Dunning basis set⁹⁰ at the LMP2 level of theory in methanol ($\epsilon = 32.63$) using the COSMO model. As a comparison with wavefunction methods, additional DFT calculations were carried out by Dr. Johannes Dieterich in a course of a collaboration.⁹¹

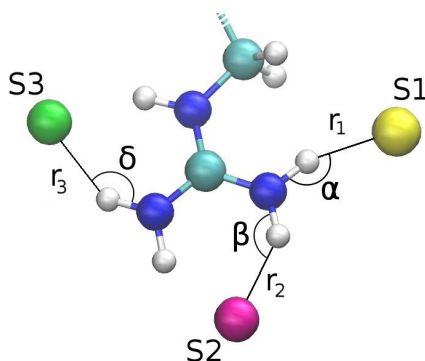


Figure 3.1: Schematic representation of the three interaction sites considered for building the protonated arginine minima. At sites S2 and S3, the anion/water species interact with two protons.

Density-fitting (DF) calculations have been used throughout,^{43,92} so one will drop the “DF”-prefix in this chapter. The density-fitting basis sets used were the JKFIT and MP2FIT for the corresponding basis.⁹³ The local MP2 calculations were carried-out using Pipek-Mezey localized orbitals.³⁴ The orbital domains were determined with the use of natural population analysis criteria, with $T_{\text{NPA}}=0.03$.³⁷ Canonical and local MP2 calculations were carried out with a development version of Molpro 2010.2.³⁸

It is important to note that for every considered molecule in this study, independent of the dielectric constant used, the domains were found to be constant. This was also verified in the case of β -peptides structures, where different conformers have been computed, the applied criteria led to the same domain definition.

3.2.2 Benchmark Results

One started by optimizing a series of small molecules. The molecules taken into account were: water, ammonia, furan, ethanol, diethyl ether, cytosine, acetyl alcohol, glycine. The selected geometrical parameters are shown in Table 3.1 for optimizations with COSMO-MP2 and COSMO-LMP2 methods using a cc-pVTZ basis set.

As it is visible in Table 3.1, the differences between canonical and local MP2 are rather small. One can note the trend is, as in gas-phase,³⁹ that LMP2 gives slightly larger bonds. This effect is due most likely to reduced basis set superposition effects, a characteristic from local methods by construction.^{44,94} However, these differences are usually around 0.001-0.004 Å. The values obtained for the angles are also quite comparable between both methods. The effect of the variation of the solvent is very small, being in

the same order of magnitude as the differences observed between MP2 and LMP2. One should note that the trends are kept, irrespective of the method.

In this first set of molecules, acetyl alcohol and glycine are quite interesting due to their intramolecular hydrogen bond. The latter is the smallest of 21 natural α -amino acids found in eukaryotes. These can serve as model systems for protein coils. Looking in more detail in the acetyl alcohol geometric parameters one can note that the solvent has a bigger impact in the internal hydrogen bond length $r(\text{H}\cdots\text{O})$ and angle $\alpha(\text{O}\cdots\text{H}-\text{O})$. In toluene the internal hydrogen bond is 0.029 Å shorter with MP2 than with LMP2. As pointed out before, this should be an effect of reduced basis set superposition errors in the local methods in comparison to the canonical ones.

Comparing the hydrogen bond length in toluene and water one observes that in toluene the bond length is 0.007 Å shorter from the canonical to the local method, while in water this difference is 0.005 Å. This shows that the solvent has a similar effect. The same holds for the angle of the hydrogen bond which is calculated to be 120.6 degrees with MP2 and 119.9 degrees with LMP2 in toluene. Changing to aqueous solvation reduces the angle by 0.4 degrees for MP2 and 0.1 degrees for LMP2, again preserving the same overall trend.

As a further test set, the structure of a few dimers of water, ammonia and formic acid were also optimized. The geometries correspondent to a minimum in toluene using COSMO-LMP2 are presented in Figure 3.2. Intermolecular degrees of freedom (DOFs) are more strongly affected by the use of continuum solvation models (CSMs) and/or local correlation approaches than internal parameters. The selected intermolecular DOFs are presented in Table 3.2. Just as observed in the case of monomers, the geometric parameters obtained for canonical and local MP2 agree very well, with slightly larger intermolecular distances in the latter structures.

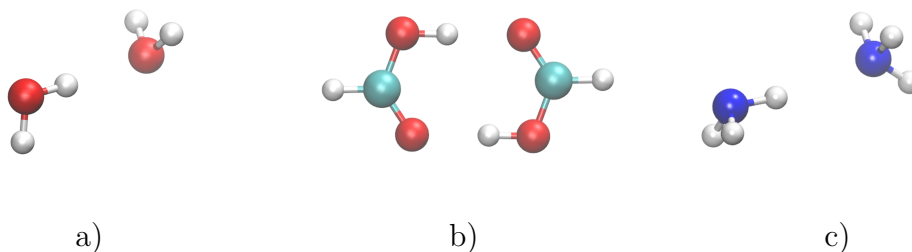


Figure 3.2: Dimer optimized geometries with COSMO-LMP2 in Toluene. a) water dimer b) ammonia dimer and c) formic acid dimer.

Table 3.1: Selected geometry parameters for different optimized structures at the COSMO-MP2 and COSMO-LMP2 levels of theory. The basis set used was cc-pVTZ throughout. The bond distances are given in Å, angles in degrees. Two solvents were considered, toluene ($\epsilon = 1.89$) and water ($\epsilon = 80.4$).

Molecule	DOF	Toluene			Water			Toluene			Water		
		MP2	LMP2	MP2	MP2	LMP2	Molecule	MP2	LMP2	MP2	MP2	LMP2	LMP2
water	r(O-H)	0.960	0.960	0.961	0.962		cytosine	1.371	1.374	1.362	1.362	1.364	1.364
	α (H-H-O)	103.5	103.4	103.1	103.0			1.223	1.223	1.238	1.238	1.238	1.238
ammonia								1.010	1.010	1.009	1.009	1.010	1.010
	r(N-H)	1.012	1.013	1.013	1.014			117.2	117.2	117.6	117.6	117.6	117.6
	α (H-N-H)	105.8	105.7	105.2	105.2			124.6	124.5	123.3	123.3	123.3	123.3
								119.9	119.9	119.7	119.7	119.7	119.7
furan	r(O-H)	1.360	1.361	1.362	1.364			12.5	12.6	3.5	3.5	4.9	4.9
	r(C-C)	1.365	1.365	1.365	1.366								
	r(C-C)	1.426	1.428	1.428	1.429								
	α (C-O-C)	106.8	106.8	106.9	106.9		acetyl alcohol	1.402	1.406	1.409	1.409	1.413	1.413
ethanol	α (O-C-H)	115.9	115.8	115.9	115.9			0.970	0.970	0.971	0.971	0.970	0.970
	α (O-C-C)	110.5	110.5	110.4	110.4			1.221	1.221	1.224	1.224	1.224	1.224
								2.000	2.029	1.993	1.993	2.024	2.024
	r(O-H)	0.961	0.962	0.963	0.963			119.2	119.5	119.0	119.0	119.3	119.3
	r(O-C)	1.426	1.428	1.430	1.433			111.4	111.7	111.2	111.2	111.5	111.5
	r(C-C)	1.510	1.514	1.510	1.514			104.2	104.6	104.2	104.2	104.7	104.7
	α (H-O-C)	107.7	107.5	107.3	107.1			120.6	119.9	120.2	120.2	119.8	119.8
	α (O-C-H)	110.7	110.6	110.3	110.3			3.2	3.2	3.2	3.2	3.4	3.4
diethyl ether	α (O-C-C)	107.5	107.6	107.8	107.9								
							glycine	1.210	1.210	1.215	1.215	1.216	1.216
	r(O-C)	1.416	1.420	1.421	1.424			1.350	1.352	1.340	1.340	1.342	1.342
	r(C-H)	1.096	1.097	1.095	1.096			0.970	0.970	0.972	0.972	0.972	0.972
	r(C-C)	1.510	1.514	1.510	1.514			1.012	1.013	1.013	1.013	1.014	1.014
	α (C-O-C)	111.5	111.5	111.3	111.3			2.279	2.281	2.291	2.291	2.292	2.292
	α (O-C-C)	108.3	108.4	108.4	108.5			123.0	123.0	123.1	123.1	123.1	123.1
	α (O-C-H)	109.8	109.8	109.6	109.6			106.1	106.0	107.0	107.0	107.0	107.0
								76.0	76.1	75.5	75.5	75.2	75.2
								62.1	63.5	66.4	66.4	67.7	67.7

Table 3.2: Selected geometric parameters for water, ammonia and formic acid dimer structures at the COSMO-MP2 and COSMO-LMP2 levels of theory. The basis set used was cc-pVTZ throughout. The bond distances are given in Å and angles in degrees. The 'd' and 'a' superscripts are used to distinguish between hydrogen bond donor and acceptor, respectively. Two solvents were considered, toluene ($\epsilon = 1.89$) and water ($\epsilon = 80.4$).

Molecule	DOF	Toluene		Water	
		MP2	LMP2	MP2	LMP2
(H ₂ O) ₂	r(O ^a ...H)	1.911	1.958	1.844	1.890
	α (O ^a ...H-O ^d)	176.0	175.4	180.0	179.4
	θ (H-O ^a O ^d -H)	124.3	124.5	123.0	123.3
(NH ₃) ₂	r(N ^a ...H)	2.244	2.297	2.202	2.257
	α (N ^a ...H-N ^d)	172.9	169.9	177.8	179.4
	θ (H-N ^a N ^d -H)	68.8	72.6	65.2	68.5
(HCO ₂ H) ₂	r(O ^a ...H)	1.653	1.656	1.656	1.717
	α (O-C=O)	126.2	125.6	125.6	125.6
	α (O ^a ...H-O ^d)	178.9	177.7	177.6	178.2

One could discuss these results in terms of variations on bond lengths and angles with the continuum, but the conclusions wouldn't be more than a repetition from what was mentioned before. It would be more interesting to advance to more usual application systems like the interaction of charged systems, as arginine with halide anions, and the conformation of peptides, in this case a small β -peptide. The latter are known to adopt a large number of secondary structures, presenting also a dependence on the solvent.

3.2.3 Amino acid Interactions

In this section one presents the application of COSMO-LMP2 to analyse specific amino acid interactions with water and halide anions. The guanidinium cation has deserved

considerable attention in the past,^{95,96} as it has been shown to serve as an effective motif for binding anions.⁹⁷ This moiety is present in the protonated arginine side chain, which can be used by proteins to dock anionic species. Up until now, quantum mechanical calculations have been mostly performed in the gas phase. Computed interaction energies of guanidinium and chloride are in the range of $450 \text{ kJ}\cdot\text{mol}^{-1}$. The study of such systems in solution is quite relevant. The guanidinium cation is present in the protonated arginine in natural proteins and enzymes. Therefore the study of such systems is highly connected to the comprehension of effects that contribute to the stability and functionality of proteins and enzymes in solution.

In the study of interacting species in solution, one is, most of the times, interested in the solvated complex energy formation. This is the difference between the energy of the solvated complex and the energy of the separate constituents, also solvated. It is expected that the latter should be significantly smaller when studying an ion-pair, since the solvation energy for the complex will be much lower than the separated species.

In order to study the systems above mentioned, one considered a series of minima around the arginines guanidinium group. The docking sites are displayed in Figure 3.1. Calculations were carried out in both gas phase and with COSMO in order to include water solvation effects. One measured the distance between the amino acid to each interacting site and also the angle formed according to the scheme represented in Figure 3.1. The contact with the acidic hydrogen from the guanidinium group is taken as reference. The results are presented in Table 3.3.

Upon the geometry optimization of arginine with chloride in position S1 in the gas phase, chloride actually deprotonates the guanidinium group, and so these results are omitted from Table 3.3. The geometries are presented in Figure 3.3. The same effect is not observed in the optimization using COSMO. This is naturally related with the role of the solvent in stabilizing the system. This shows the importance of the solvent effects in the geometry, since a study of complex formation energy of this system would be impossible to perform with gas phase geometries. The same effect was also observed for bromide in the same position, although with to a lesser extension, since the proton abstraction was not verified. Nonetheless the distance between the bromide and the proton is shorter in the gas phase by 0.4 \AA , which is a sign of strong attraction. The same trend was also observed for other interactions sites with chloride and bromide. The distances in gas phase are significantly shorter ($0.3\text{-}0.5 \text{ \AA}$). In these sites no deprotonation was observed.

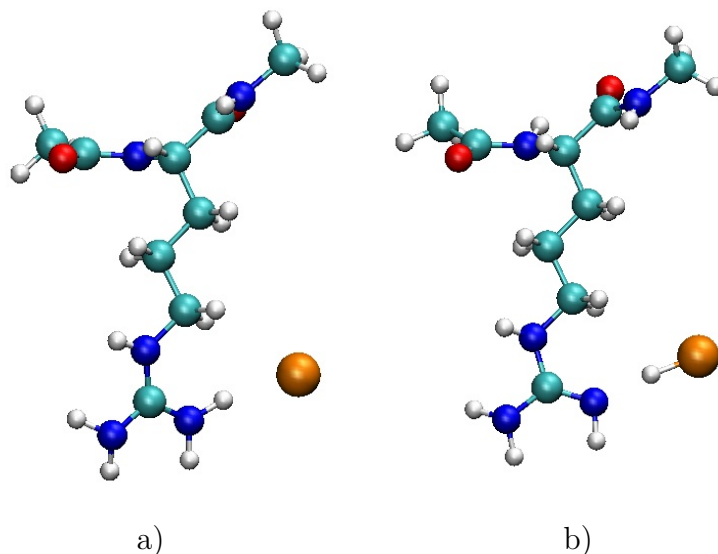


Figure 3.3: Optimized structures of protonated arginine and chloride in position S1: a) in water ($\epsilon=80.4$) and b) gas phase.

For the arginine/water system, the distance was measured between the oxygen and the proton of the guanidinium group. For this case the variations in the distance between solution and gas phase are one order of magnitude smaller (0.02-0.04 Å) than the ones registered in the case of the anions. This small variation is a result of the fact that the water is not charged. In general, one can observe that gas phase geometries optimizations for the present cases tend to overestimate the interaction, as expected.

The interaction energies for the studied systems were also computed, these results are presented in Table 3.3 in solution and in the gas phase. These were computed by single points calculations at the LMP2/aug-cc-pVTZ level of theory. The internal geometry of the separated components of each system were the same as in the complex optimization. The values obtained for solution are quite different from gas phase. This is related to the fact that the geometries that resulted from the structure optimization are different and for solution simulation one also takes into account the difference in the solvation energy of each of the species.

The interaction energies in the docking site S1 are weaker than in the others. This is related to the fact that in the latter only one interacting proton is available. The difference to the other sites is quite constant with differences around 10 kJ·mol⁻¹. It is important to notice that chloride and bromide present larger stabilization energies

than water. This goes towards the idea that arginine can interact with anions and that these can compete with water to interact with the guanidinium moiety of arginine. The energy differences in solution are around 7 to 15 kJ·mol⁻¹ depending on the anion and the docking site. Much bigger differences are registered for gas phase. Nevertheless the latter are quite comparable with calculations with only the guanidinium moiety. Rozas *et al.*⁹⁶ reported an interaction energy between -437.7 and -449.9 kJ·mol⁻¹ at site S2 for the chloride anion. One computed an interaction of -476.0 kJ·mol⁻¹ which is somewhat closer to their CBS-QB3⁹⁸ value of -473.8 kJ·mol⁻¹. The anions interaction energies show the largest difference to the values in solution. This is connected to the fact that a neutral complex is formed, with a lower solvation energy compared to what one obtains for the separated ions. In the case of water, the complex formed is positively charged.

Table 3.3: Geometric parameters (distances in Å, angles in degrees) for the interacting anion/molecule and the acidic proton in the optimized arginine systems. The calculations in the gas phase were carried out with LMP2/cc-pVDZ, calculations in water with COSMO-LMP2/cc-pVDZ ($\epsilon = 80.4$). The energy difference ΔE (in kJ·mol⁻¹) between the complex and the isolated species has been computed with the aug-cc-pVTZ basis set.

		Water			Gas phase		
		Distance	Angle	ΔE	Distance	Angle	ΔE
S1	Cl ⁻	2.132	166.4	-26.3	—	—	—
	Br ⁻	2.382	163.4	-19.8	1.964	171.1	-430.1
	H ₂ O	1.864	168.3	-11.6	1.844	171.3	-51.9
S2	Cl ⁻	2.346	153.0	-35.9	1.917	156.4	-476.0
	Br ⁻	2.496	156.0	-28.5	2.130	156.5	-443.6
	H ₂ O	1.971	149.6	-21.9	2.003	149.1	-65.7
S3	Cl ⁻	2.175	160.0	-34.2	1.865	162.0	-483.5
	Br ⁻	2.370	162.0	-30.7	2.075	164.2	-481.7
	H ₂ O	1.938	151.8	-20.9	1.973	149.9	-100.4

The interaction sites S2 and S3 seem to be more energetically favourable when com-

pared to S1. This maybe due to the fact that in these positions the species interact with two acidic hydrogens. Another interesting observation is that in site S3, contrary to the other sites, the chloride and bromide present a very similar energy, with differences smaller than $3.5 \cdot \text{mol}^{-1}$, while one would expect a stronger interaction for chloride. This could be related to a poor sampling during the complexes optimization or to shortcomings in the use of CSM for this particular application. Maybe a more accurate description of the system using explicit solvent molecules would be needed. One also optimized the previous systems using canonical MP2 with COSMO and compared the distances with the local approximation, shown in Figure 3.4. The solid line represents the case where the distances obtained at MP2 and LMP2 levels would be strictly the same.

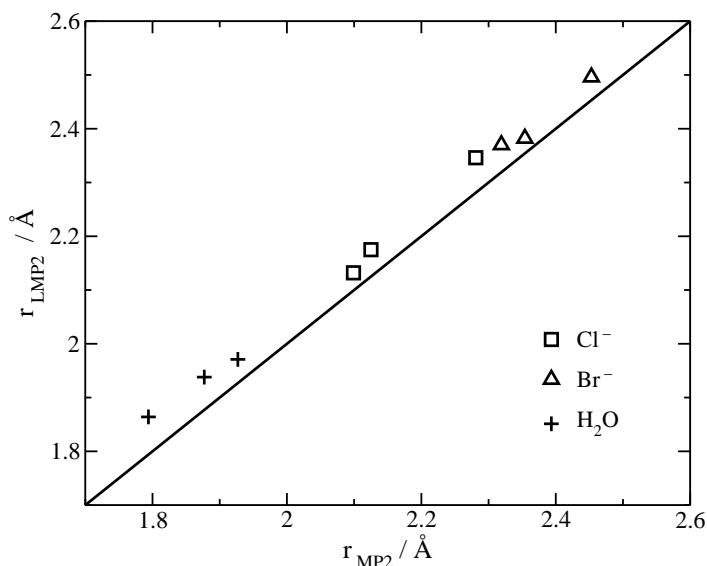


Figure 3.4: Shortest H-X ($X=\text{Cl}^-$, Br^- , H_2O) distances, in \AA , for the optimized structures using COSMO-MP2 and COSMO-LMP2 using cc-pVDZ basis set.

As expected, the structures optimized with LMP2 present slightly longer distances, with deviations between 0.03 to 0.07 \AA when compared with MP2. Nonetheless the LMP2 distances follow the same trend as the MP2.

3.2.4 β -Peptide Conformers

In this section, we take a look into β -peptide conformations. β -peptides conformational studies will be discussed in more detail in Chapter 5. These oligomers are identified ac-

cording to their intramolecular hydrogen bonds pattern in the backbone.⁹⁹ The majority of first principle calculations done to date for these kind of systems has been achieved with DFT methods. The latter are known to lack the description of dispersion forces. In order to overcome this flaw, empirical corrections were recently proposed by Grimme *et al.*⁵³ Nonetheless the latter are defined independently of the surroundings, being their contribution to the energy the same for gas phase and for solution. In order to obtain a consistent treatment of the interactions within a CSM description, a dependence on the external potential should be included. In this context one will use *cis*-ACPC tetrapeptides in the above mentioned conformations, using the proposed method. This will test the applicability of the latter in the study of large systems. The *cis*-ACPC tetrapeptides molecules are composed of 74 atoms. Upon geometry optimization with starting hand-generated structures, two conformations were obtained, a sheet and a 14-helix, present in Figure 3.5. The sheet conformer is very similar to an unfolded protein structure. The 14-helix is a coil like structure, which is characterized according to a specific hydrogen bond pattern in the backbone. The latter occurs between the amides hydrogen atom from the i th monomer and the carbonyl group from the $(i+2)$ th monomer. The smallest possible oligomer to present such conformation is a tetrapeptide. In experimental β -peptides synthesis one of the commonly solvents used is methanol. This is the reason for its use in this simulation.

Upon optimization, the folded tetrapeptide structure does not correspond a hundred percent to a 14-helix. From the two observed turns only one presents a 14-helix pattern, the second shows a 10-helix loop contact. This is related to the fact that for a small oligomer system, like a tetrapeptide, in a 14-helix structure the ending groups have also some influence on the terminal hydrogen bond pattern. The backbone in the terminal monomers, when compared to inner monomers, in the peptide, present a higher freedom of movements. This can lead to the rearrangement into other hydrogen patterns. The sheet like formation presents a conformation comparable to the one observed for larger peptide chains.¹⁰⁰ In a published work with other authors⁹¹ one presents results from the conformational stability of the two oligomers using LMP2 and density functional calculations. The results are presented in Table 3.4.

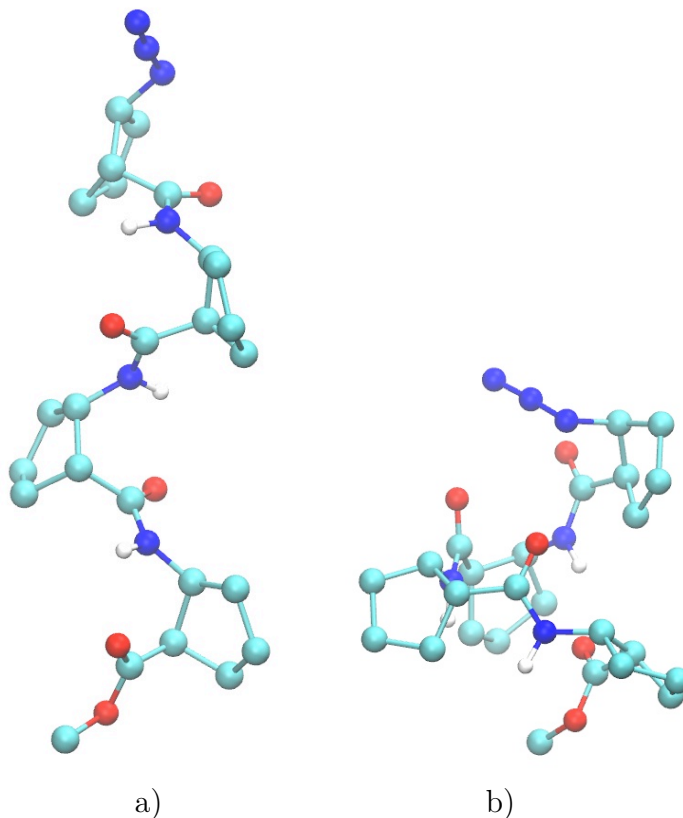


Figure 3.5: Optimized structures of the *cis*-ACPC tetrapeptide conformers with COSMO-LMP2/cc-pVDZ in methanol ($\epsilon = 32.63$): a) β -sheet, b) 14-helix conformation. The allylic hydrogens were removed for a better visualization.

Table 3.4: Energy difference [$\Delta E = E(\text{sheet}) - E(\text{helix})$] between the two conformers of the *cis*-ACPC tetrapeptide (in $\text{kJ}\cdot\text{mol}^{-1}$). Geometry optimizations and single points were computed at different levels of theory.

Geometry	Energy	ΔE
LMP2/cc-pVDZ	LMP2/aug-cc-pVTZ	24.6
	BP86/def2-TZVP	-41.4
	B3LYP/def2-TZVP	-37.6
	BP86-D3/def2-TZVP	12.9
	B3LYP-D3/def2-TZVP	9.2
BP86/def2-SVP	BP86/def2-TZVP	-36.4
BP86-D3/def2-SVP	BP86-D3/def2-TZVP	60.3

cis-ACPC presents no aromatic side chain, and one would expect that LMP2/aug-cc-pVTZ should deliver close to converged results for the conformational energy as is observed for the case of α -peptides.¹⁰¹ For the two geometries considered, LMP2 favours the helix conformer by 24.6 kJ/mol. The same trend is observed for DFT-D3. In general the dispersion correction seems to be a determinant factor, since pure DFT functionals favour the sheet like conformation. Unlike in the latter, in the helix conformation one would expect that dispersion forces would play a more dominate role. The importance of dispersion corrections for this system can be notice when comparing the optimization and single points at the DFT level with and without dispersion corrections.

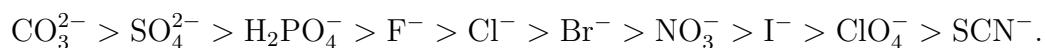
The COSMO-LMP2 optimizations were carried out with the cc-pVDZ basis set. A full Hartree-Fock calculation plus LMP2 and gradient calculation (with density fitting approximations) took about 4.3 hours.⁹¹ In the latter more than half of the time was taken by the LMP2 part with the gradient.⁹¹ The use of COSMO was found to add small time variations. The latter is more significant in the HF part. This is due to the inclusion of COSMO in each HF cycle. The gradient take only an additional 7.2 minutes to compute. The use of COSMO did not demonstrate to affect the memory requirements.⁹¹ The results show that such an approach could be effectively applied to the studies of conformational properties of a small oligopeptide composed of 74 atoms, at a high level of theory.⁹¹

4 Specific ion-amino acid interactions

4.1 Overview

The role of specific ion effects on the structure of water and proteins in ionic solutions is a key aspect to the understanding of different processes in chemistry and biology. The molecular mechanisms underlying the stability of proteins in distinct ionic solutions are partially hidden in the interplay of these specific ion interactions.

Such effects are known to follow the trend of the so called Hofmeister series.¹⁰² In the later years of the nineteenth century Hofmeister and his co-workers observed that some salts present a more precipitant character (salting-out) on the egg white protein, while others maintained the protein in solution (salting-in).¹⁰³ He ordered these salts in a series according to their ability to produce the salting-out, known in our days as the Hofmeister series.¹⁰⁴ The anionic Hofmeister series ordered according to the ability to originate salting-out is¹⁰²



In the previous chloride normally is the delimiter between the anions that cause salting-out (left-hand side) and salting-in (right-hand side). The formers are known to increase the protein stability leading to a consequent decrease of their solubility. The anions on the right-hand side of chloride present the opposite properties of the latter.

Originally, the effects that lead to the anions ordering in the series were described in terms of ion-solvent interactions.¹⁰² Anions on the left of chloride are called kosmotropes (structure makers) while anions on the right are called chaotropes (structure breakers).¹⁰⁵ Kosmotropic species are high hydrated ions that are able to strengthen the hydrogen-bond network of several layers of water molecules (water makers) and can even remove waters from the protein solvation shell inducing their precipitation.^{105,106} On the other hand, chaotropic species are weakly hydrated ions that eventually interrupt the

hydrogen-bond network.^{105,106}

The concepts above mentioned are related to the ability of the anions to reshape the hydrogen bond pattern from their solvation shell into the bulk (so described as a long-range effect). However, these long-range effects are still a matter of debate.^{105,107} Recent experiments have cast serious doubts on these concepts. Bakker and co-workers studied the magnesium perchlorate and sodium perchlorate and sulphate in aqueous solution using femtosecond pump-probe spectroscopy. They observed that the ions did not have any influence on the water dynamics outside their first solvation shell.¹ Later in 2004, Pielak and co-workers using pressure perturbation calorimetry studied the impact of protein stabilizing and destabilizing solutes on the water structure and found no correlation.² More recently Saykally *et al.* studied the water OH vibration through Raman spectroscopy in the presence of potassium halide salts and also did not find any evidence in support of the previous nomenclature.³ Suggestions have been made to change the classification in terms of weak and heavy ions.¹⁰⁸ These observations do not mean that water-ion interactions are unimportant, only insufficient do describe the ion position in the Hofmeister series. Therefore, in a molecular description of protein solubility in the presence of different salts one should also take in account protein-water, protein-ion interactions.¹⁰² It should be emphasize that the target of this work is not to clarify the concepts of structure makers and structure breakers. Instead, one is focused in to study the nature of peptides-ion interactions in solution and their role in protein stability.

In this section we take a closer look at the interaction of specific anions with several amino acids. In order to address the ionic affinity towards different amino acids residues we have focused on monovalent halide anions. The latter are positioned in the Hofmeister series in the following relative order $F^- > Cl^- > Br^- > I^-$. Iodide is a protein denaturant promoting protein solubility (salting-in) and on the other end, fluoride increases protein stability and decreases protein solubility (salting-out).

4.2 Anion Specific Effects

In order to study the specificity of ion-amino acid interactions, one combined five amino acid models and four different halide anions. The models considered were the neutral histidine and the protonated forms of histidine, arginine, lysine and a model system

for the backbone composed of alanine. These charged systems were chosen taking into account a high variety of possible interaction sites with anions. Arginine, histidine and lysine are known to exist in the cationic form, being composed of the guanidinium, imidazolium and ammonium terminal group respectively. The halides anions used were fluoride, chloride, bromide and iodide. The choice of this anionic series is related with the fact that it allows the study of ion-amino acid interactions within a series from small hard anion like fluoride to a hard soft anion like iodide. Another reason its also their position in the Hofmeister salting-out series. The interaction of these four amino acid systems with the same halide anions were also considered in molecular dynamics studies by Heyda *et al.*¹⁰⁸

4.2.1 Computational details

All calculations, unless otherwise noted, were performed with the LMP2 method, where solvation effects were introduced through the use of the COSMO model, as described in section 3.1. The dielectric constant (ϵ) used to describe the aqueous environment was 80.10 with the default optimized atom radii (H = 1.30, C = 2.00, N = 1.83, O = 1.72, F = 1.72, Cl = 2.05, Br = 2.16 and I = 2.32 Å). The amino acid models were acetylated in the N-terminus and methylaminated at the C-terminus. This was carried out in order to prevent the influence of charged terminal groups and to obtain a more protein-like description of the backbone. With these approximations one can simulate intermediate amino acids side-chains on the protein surface. Geometry optimizations were carried out at the COSMO-B3LYP/6-31G* level of theory,⁶ with different torsion angles of the backbone section. The most stable structures were taken for further calculations. In order to characterize the potential energy surface of the interaction of the anion with the amino acid models, the energy was calculated by placing the anion in several points of an automatically generated grid, while the amino acids were kept fixed.

For each of the five studied systems under study, a point grid with a 0.2 Å mesh was built. The amino acids grid was aligned according to the inertia tensor of the side-chain, computed from the gamma carbon onward (C_γ) and excluding hydrogen atoms. In the case of the backbone model all atoms were included with exception of the methyl side-chain of the alanine.

In the description of the PES one started by calculating the LMP2 energy and the domains at a large non-interacting distance (100 Å). Then the LMP2 energy was recom-

puted in each point of the grid, placing the ions in a certain position using the saved domains for consistency.⁹⁴ The interaction energy was taken as the difference between the LMP2 value at the grid point and the non-interacting configuration. The grid points next to the model system were rejected according to scaled atomic van der Waals radii (with a factor of 0.6) and on the basis of an energy cutoff (interaction energies superior to 50 kJ·mol⁻¹ were neglected). For the PES, the interpolation of the surface points was performed through the use of bicubic splines using MATLAB.

The basis set used were correlation-consistent with triple ζ valence Dunning basis sets (cc-pVTZ, up to p functions for hydrogen atoms and f for the heavier atoms), with added diffuse s and p functions to non-hydrogen atoms. For bromide and iodide, Stuttgart-Dresden pseudopotentials were used (EPC10MDF for Br⁻ and EPC28MDF for I⁻), together with the cc-pVTZ-PP basis set (up to f functions) and aug-cc-pVTZ-PP s and p diffuse functions.^{109,110} Density-fitting approximations were used with cc-pVTZ/JKFIT and cc-pVTZ/MP2FIT,⁹³ except for I⁻ for which the QZVPP/JKFIT¹¹¹ and QZVPP/MP2FIT¹¹² were applied. The basis sets used were a compromise between accuracy and computational cost since each PES is composed of several thousand points.

Single point calculations in each point of the grid showed a maximum error of 3 kJ·mol⁻¹ when compared with full aug-cc-pVTZ and aug-cc-pVTZ-PP basis set. The error introduced by using density-fitting approximations were found to be negligible, with deviations inferior to 0.1 kJ·mol⁻¹. The LMP2 calculations were carried-out using Pipek-Mezey localized orbitals.³⁴ The orbital domains were determined with the use of natural population analysis criteria, with $T_{\text{NPA}}=0.03$.³⁷ All calculations were performed with a development version of Molpro 2009.2.³⁸

4.2.2 Results and Discussion

Anion Specific Effects

We will start by comparing the obtained energy surfaces of different amino acids in the presence of the same anion. Points with energy above 50 kJ·mol⁻¹ were excluded from all plots. In this first part one is only interested in the interaction of the anion with the amino acid side chain, therefore ion/backbone interactions will be discussed and presented far ahead. In Figure 4.1, the results with four amino acid model systems interacting with iodide are presented.

Upon inspecting the PES one finds, as expected, that the minima are found next to

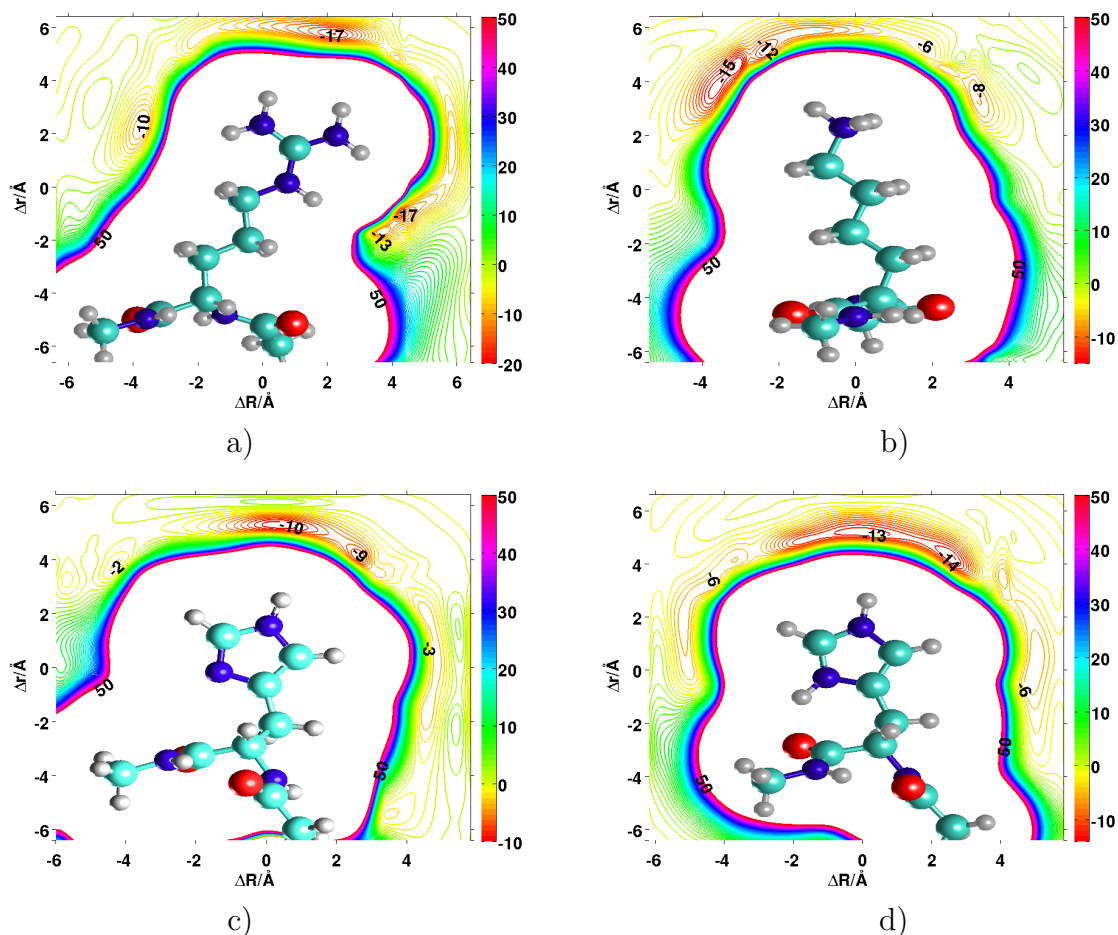


Figure 4.1: Potential energy surfaces (energy in $\text{kJ}\cdot\text{mol}^{-1}$) for different amino acid models interacting with the I^- anion. a) protonated arginine, b) protonated lysine, c) histidine and d) protonated histidine.

the positively charged sites, such as guanidinium, imidazolium and ammonium groups. Considering the minima depth, one can order the interaction of the iodide with the moieties from the most to the less stable as follows: guanidinium > ammonium > imidazolium.

One important piece of information that is left out, due to this two dimensional approach, is the study of anion interactions with the histidine ring and with the guanidinium group along the plain normal axis. There could be favourable interactions between the π -system and the anion since the former holds a positive charge. Calculations along the normal axis of histidine and arginine were also carried out and are later presented and discussed.

Figure 4.2 depicts a series of PES for the fluoride anion. These results should be considered with caution since the latter is a much stronger base than the other anions. A proton abstraction of the ammonium group in lysine moiety could occur. Nevertheless, since in the used approach the amino acid geometry is kept fixed, this is not possible. The interaction energy could be in any case overestimated. Nonetheless, these results are important in order to have a more complete trend of the anions behaviour. As expected in the case of fluoride, the minima depth are significantly more pronounced when compared with the minima found for iodide. One can observe a stronger interaction with the ammonium group, weaker with guanidinium. The minima stability can be ordered as follows: ammonium > imidazolium > guanidinium. The guanidinium group, however, in opposition to the charged histidine and lysine holds more comparable minima. This is a reflex of the electronic delocalization in the protonated group. One should note that in protonated histidine only a single acidic hydrogen atom is accessible to interact with the anion in this specific geometry. In the case of the other acidic hydrogen, this presents an intrinsic intramolecular interaction with the amide carbonyl group of the backbone. Nevertheless, this second acidic hydrogen is responsible for a deformation of the PES when comparing the neutral and protonated histidine. It is also important to refer that in Figure 4.2 d), the hydrogen of the terminal methyl is shown to be overlapping with the potential energy surface, when in fact it is above the plane of the grid.

Comparing the behaviour of the iodide and fluoride with the neutral and protonated histidine systems, one can notice that in the case of iodide, the presence of an extra hydrogen in the amino acid has little influence on the energy profile. Two main minima are observed. The first minimum depth changes from -10 to -13 kJ·mol⁻¹ and the second changes from -9 to -14 kJ·mol⁻¹. The profiles in the presence of fluoride show much larger differences. The more dominant minimum changes from -32 to -50 kJ·mol⁻¹, a second less intense minimum changes from -5 to -18 kJ·mol⁻¹. This results reveals how dominant electrostatic effects are in these interactions. It is interesting to notice that the more significant minimum in the neutral histidine is comparable to the values found for the protonated arginine.

Considering these results one could try to compare them with the ones obtained by Heyda *et al.*¹⁰⁸ with molecular dynamics for the same amino acid systems (although geometries differ). The latter calculation were carried out with a polarized and non-polarizable versions of the parm99 force field. In the former the POL3 water model was used, while in the latter case the SPCE water model was used. The latter authors

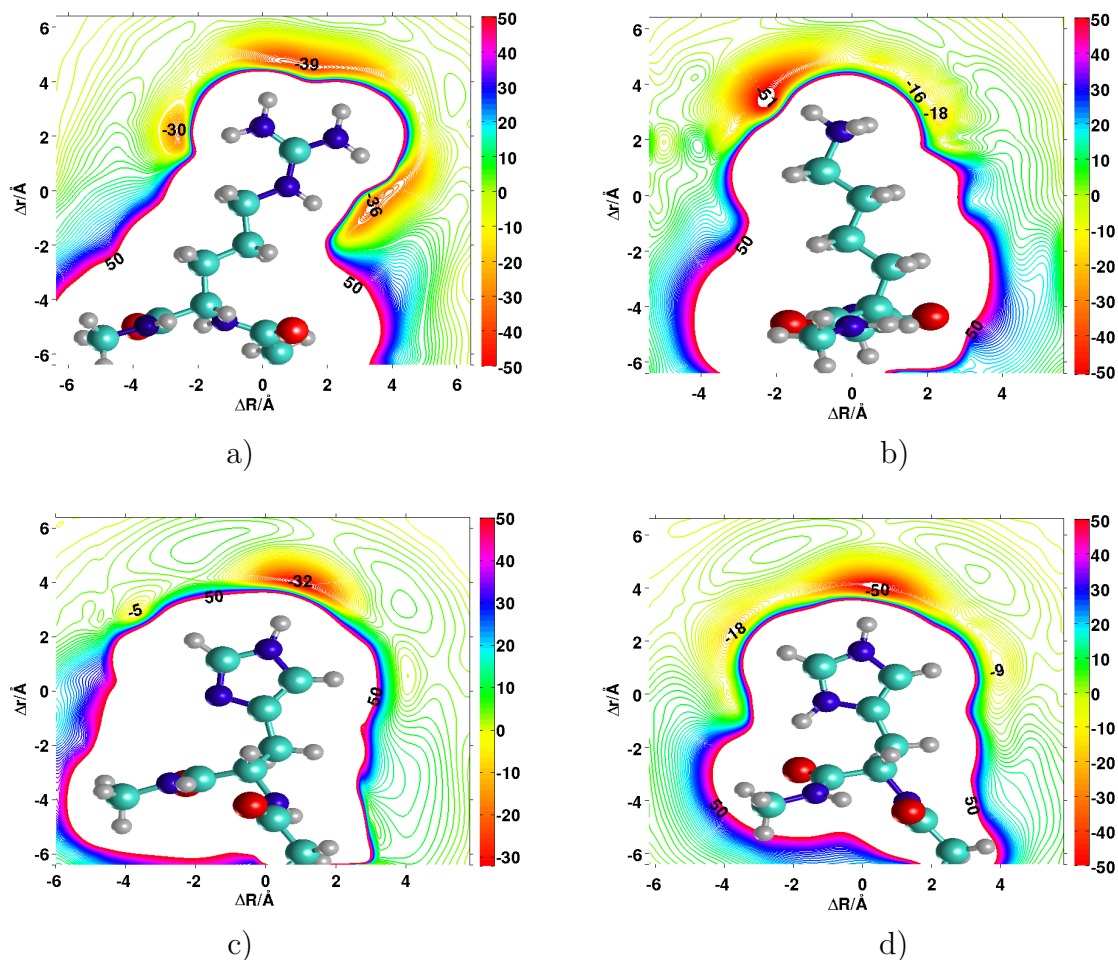


Figure 4.2: Potential energy surfaces (energy in kJ·mol⁻¹) for different amino acid models interacting with the F^- anion. a) protonated arginine, b) protonated lysine, c) histidine and d) protonated histidine.

verified that with the first version of the force field, all anions exhibit a high affinity for the charged group of the amino acids. With the use of a polarized force field one could verified that larger anions presented a more diffused density distribution to more hydrophobic regions of the side chains of the amino acids.

The authors concluded that small anions like fluoride exhibit a strong affinity for positively charged groups in the order guanidinium > imidazolium > ammonium. The specific ordering is divergent with the one obtained in this work. This might be due to the fact that the positive charged groups are constrained to the plane and therefore cannot be mapped exactly onto the MD results.

Specific Arginine-Ion Interactions

The guanidinium group in protonated arginine depicts a less dissimilar minima, among the different ionic side chains studied. This might favour the observation of large residence times and density distributions of fluoride near the latter. This is due to the fact that all the hydrogens from the guanidinium group present a high acidic character.

One should consider not only the minima depth but also the overall shape of the potential. In a second set of results, interaction between the protonated arginine and the anion series is investigated. This allows one to observe specific differences between the ion species, following down the periodic table. These results are shown in Figure 4.3.

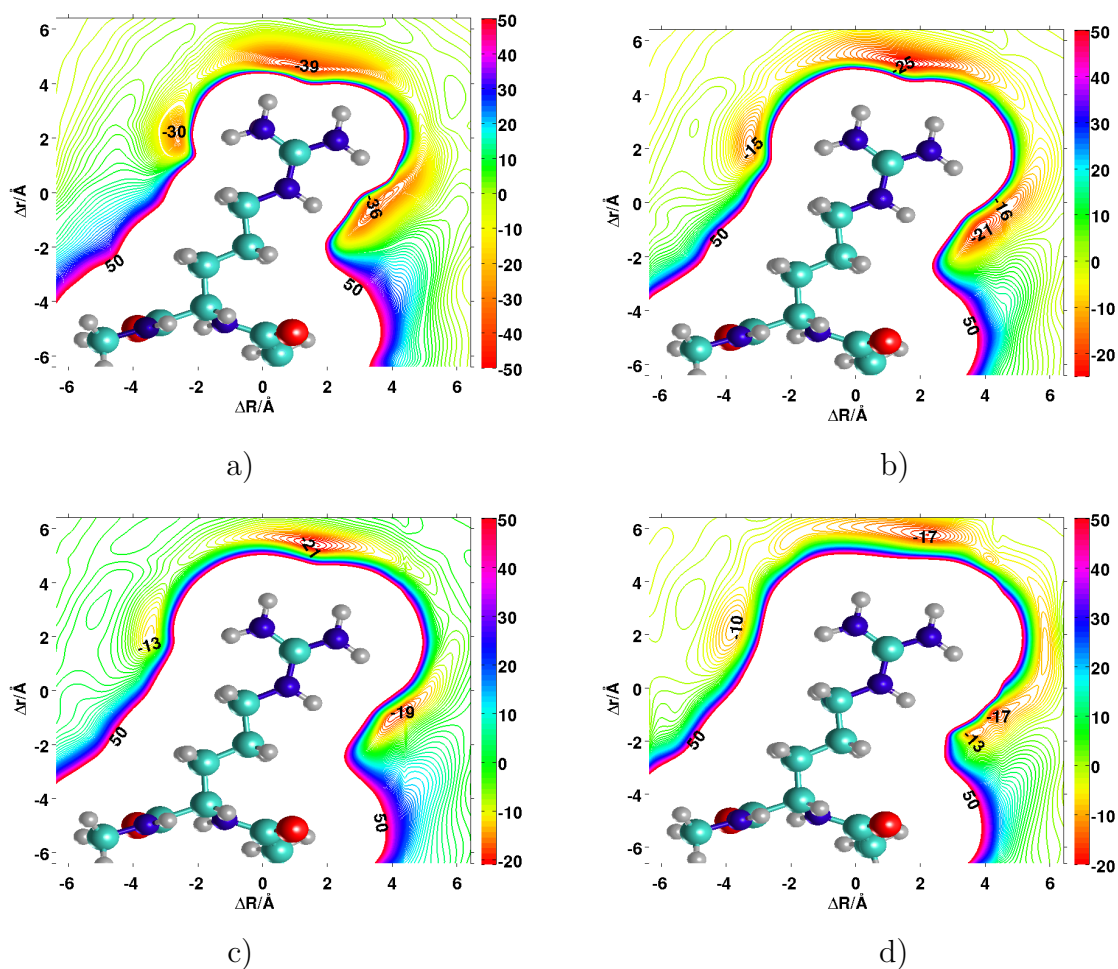


Figure 4.3: Potential energy surfaces (energy in $\text{kJ}\cdot\text{mol}^{-1}$) for different anions interacting with the protonated arginine model. a) F^- , b) Cl^- , c) Br^- , d) I^- .

Comparing the four potential energy surfaces, one observes that the location and the number of minima only changes slightly. The location is mainly influenced by the size of the anion, with fluoride showing the closest minima to the amino acid structure. The most marking difference is in the depth of the minima. Fluoride presents the strongest interaction with the moiety, with the minima energy varying between -30 to -39 kJ·mol⁻¹. These minima become consistently smaller for larger anions. For iodide they vary between -10 and -17 kJ·mol⁻¹. This is expected since fluoride will have much stronger electrostatic interactions with the charge amino acid. Iodide on the other hand, due to its very diffused electro cloud, interacts less with the positive charge in guanidinium. The profiles are slightly asymmetric around the arginine guanidinium group, mostly due to the in-plane orientation of the N-H group next to the C_δ. For all the four cases three main minima regions can be observed.

Moving farther away from the charged group, one notices further differences between the anions. It is visible that the iodide and bromide PES are much flatter than for the remaining two. For fluoride and chloride the minima are very sharp and present a repulsive potential near the non-charged groups. The difference in the shape of the potential between the several anions leads to believe that small anions present a higher residence time near the charged group, while larger anions present smaller differences in the residence time between hydrophobic and hydrophilic sites of the amino acids. One could order the anions towards the guanidinium group according to their stability as fluoride > chloride > bromide > iodide.

Comparing these results with the ones in Chapter 3.2 at the S1 docking site, the distance between chloride and the proton is 2.132 Å compared with 2.192 Å in the PES presented in Figure 4.3. The same trend can also be observed for the case of bromide in S1 2.382 Å compared with 2.392 Å. One should take in account that the basis set used in the results from Chapter 3.2 is not the same and in the present study and the whole system is relaxed (the anion is not constrained to a grid). Nonetheless the two approaches are in good agreement.

A comparison between the PES for the protonated arginine interacting with the four halide anions in this study with the density distribution from the same interacting system from Heyda *et al.*¹⁰⁸ is presented in Figure 4.4 It is very interesting that the two different approaches are in quite good agreement. The localization of the minima from our PES coincide with the locations of high anion density observed by Heyda and coworkers. This leads to believe that our approach seems suitable for the study of such systems.

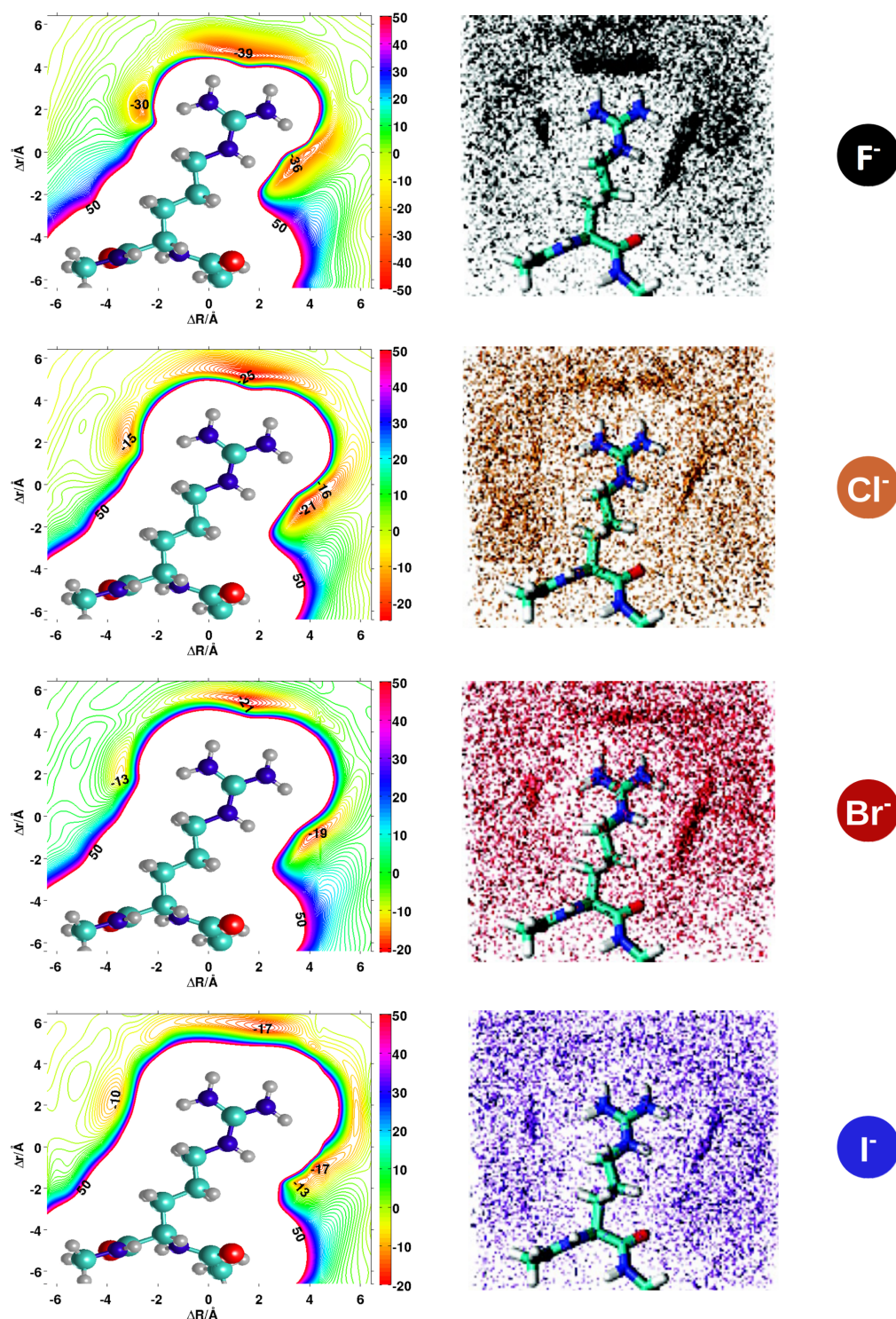


Figure 4.4: Comparison between the PES resultant from the interaction between protonated arginine and the four considered anions in this study (right) and the anions density distribution around the same amino acid (left), resultant from a MD simulations with polarizable parm99 force field. The pictures from the latter were adapted with permission from Reference [108]. Copyright 2009 American Chemical Society.

Backbone Interactions

In the previous section, the amino acids used are models systems where preferential anions interactions may occur. Nonetheless, due to the adopted approach, it was not possible until this point to retrieve information about anion interactions with the backbone. These interactions can also be a determinant factor for Hofmeister-type effects. With this in mind, one built a backbone model with alanine and computed the PES for two extremes of the chosen anion series, for fluoride and iodide. The results are presented in Figure 4.5. It is important to notice that the methyl group from alanine is above the plane of the grid, with only the amide groups of the backbone and the used caps accessible to interact with the anions.

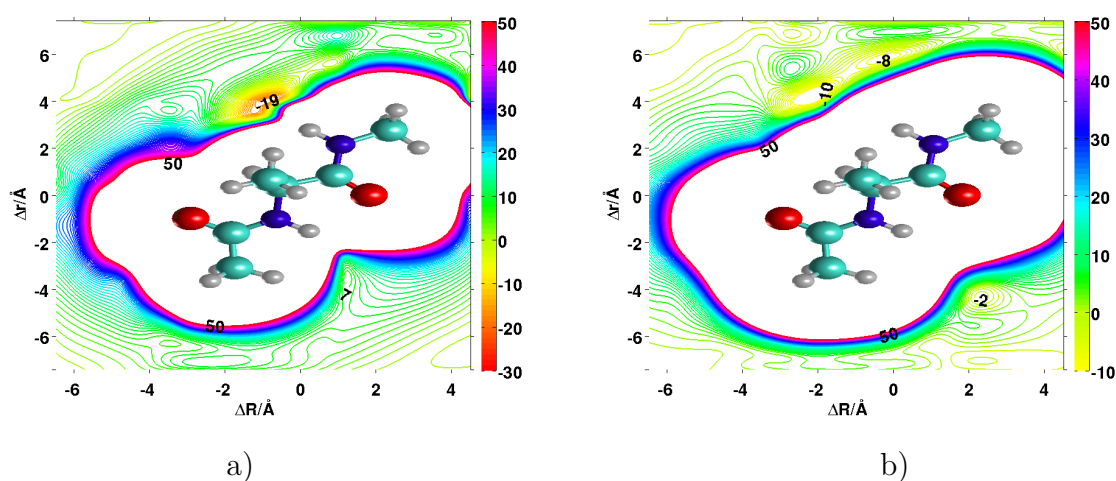


Figure 4.5: Potential energy surfaces (energy in $\text{kJ}\cdot\text{mol}^{-1}$) for fluoride (a) and iodide (b) interacting with a backbone model.

The PES shows an interaction site next to the amide N-H site. This is predictable, since the latter is expected to bear a positive charge. When comparing the potential plots for the two anions, one observes that the surface presents a much flatter profile. Interestingly, the PES for iodide depicts a less intense minimum near the second N-H with $-2 \text{ kJ}\cdot\text{mol}^{-1}$, which has a carbonyl group pointing in the same direction, whereas is absent for fluoride. This is a symptom of concomitant weak electrostatic interaction and stronger dispersion forces that counteracting the repulsion towards the carbonyl group. In the case of fluoride, one would expect a favourable interaction in cases where the carbonyl group is tilted out of the plane.

The rest of the surface is featureless. One further the repulsion of the anions towards the carbonyl group, due to the excess charge located at the oxygen.

Dispersion Contributions

In order to investigate the specific contribution of dispersion forces, one made use of the local correlation treatment to single out these interactions. In this section one present results where one subtract the dispersion contribution to the total PES. A comparison between PES of the interaction of fluoride and iodide with protonated histidine with and without dispersion forces are shown in Figure 4.6.

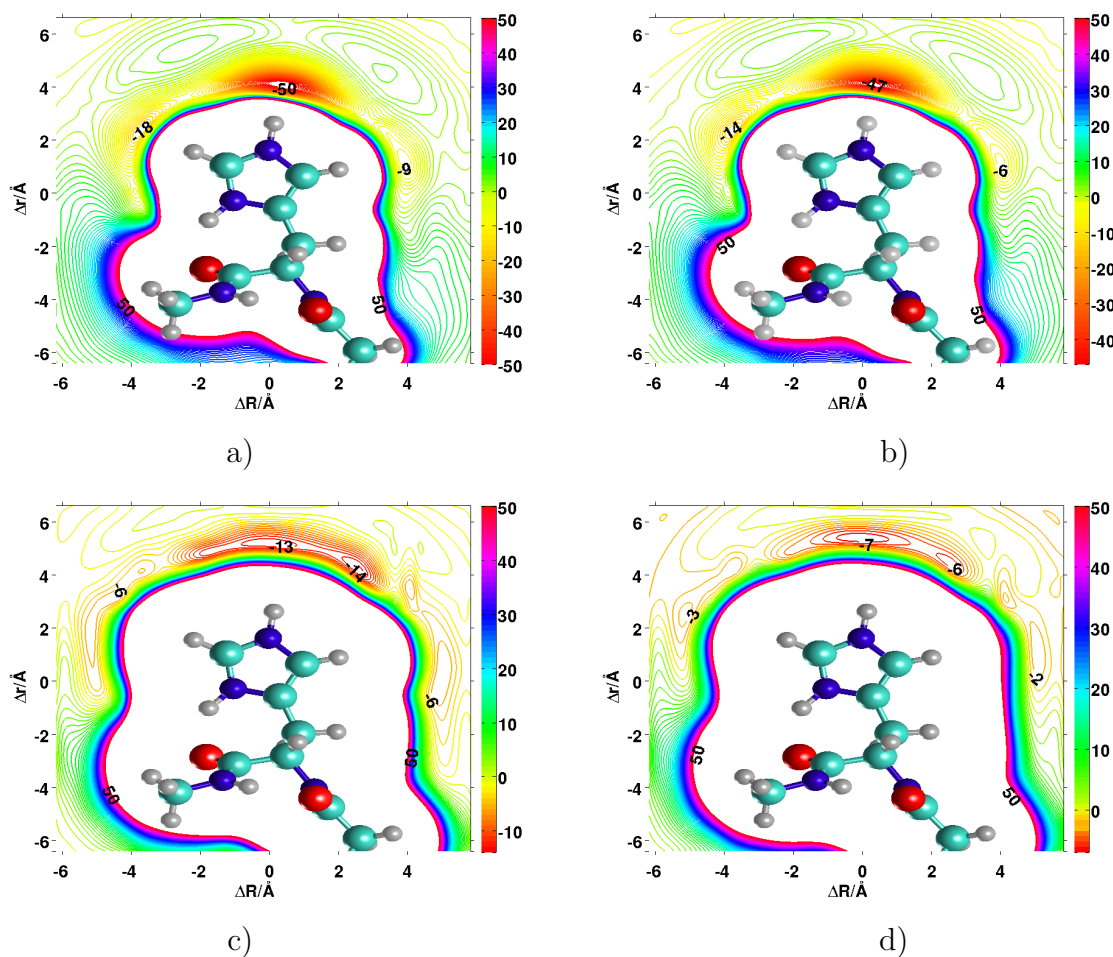


Figure 4.6: Potential energy surfaces (energy in $\text{kJ}\cdot\text{mol}^{-1}$) for the protonated histidine system with a) fluoride and c) iodide. The dispersion forces have been subtracted in b) for fluoride and d) for iodide.

One can observe that the main interactions occur around the histidine ring, as expected. The minima in the PES preserve their location upon subtraction of the dispersion forces contribution. One of the most interesting facts to note is that the dispersion contribution to the minima depth in the case of fluoride is about half of the one registered for iodide. Thus, the interaction is decreased by 3-4 kJ·mol⁻¹ for fluoride and 6-8 kJ·mol⁻¹ for iodide. The relative effect is much larger in the case of iodide than for fluoride. In the case of the first, dispersion contributes with 50 % to the more intense minima, whereas only 5 % to the latter. This difference further stresses the disparity between the electrostatics and dispersion contributions for fluoride and iodide. It also explains the relative affinity of large anions, like iodide, for apolar hydrophobic regions, in contrast to small anions like fluoride, which interacts almost exclusively with charged groups of the amino acids. This conclusion goes towards others works that address specific anion effects.¹⁰⁶

A well-known example is the molecular interpretation of the surface tension enhancement of the ionic aqueous solutions relative to neat water.^{113,114} Thus similar reasons have been pointed out to explain the affinity of large anions for both air-water interface and other hydrophobic interfaces.

Anion- π Interactions

In the previous sections one has discussed the anion-amino acid interaction in terms of 2-D potential energy surfaces, where anions are placed on an arbitrary plane. In the case of lysine, the rotation of the side-chain would lead to a different PES, but new type of interactions are to be expected. In the case of histidine and arginine, however, one should also consider the π -system and how the anions may interact from other orientations. Anion- π interactions have become a topic of interest for theoreticians and experimentalists alike.^{115,116} Similar systems have been the subject of several recent theoretical studies, but mostly without including solvation effects.¹¹⁷⁻¹²² In this case, electrostatics dominate the energy profile. However, embedded in a polar environment, the relative effect of the latter contributions will decrease. Smaller absolute contributions, like dispersion, can be particularly significant in establishing the difference between the different anions. With this purpose in mind, one carried out 1-D calculations, placing each anion along the normal axis of the histidine's ring and of the guanidinium group at varying distances. Just like the previous calculations, one started by computing the anion-amino

acid system at a large non-interacting distance, saving the domains information and then using the latter in computing each point along the normal axis. In this way a consistent description of the virtual space is used.

In Figure 4.7 one has plotted the potential energy surfaces for neutral histidine model and the four anions. One can observe that only in the case of bromide and iodide

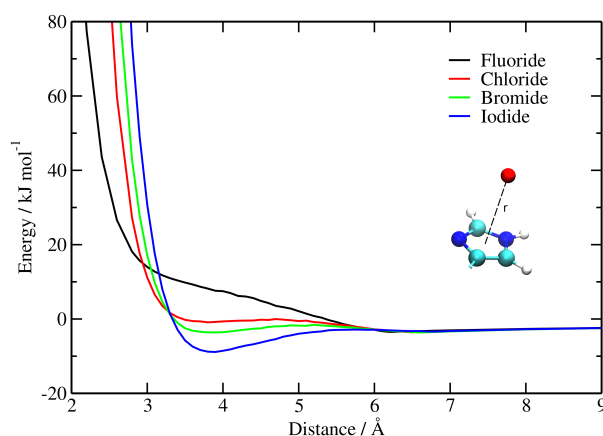


Figure 4.7: Potential energy curve (in $\text{kJ}\cdot\text{mol}^{-1}$) with COSMO-LMP2 for the neutral histidine system in combination with the anions F^- , Cl^- , Br^- and I^- , relative to the distance r along the ring normal axis.

significant minima are found, -3.6 and $-8.9 \text{ kJ}\cdot\text{mol}^{-1}$ in depth, respectively. The minimum found for iodide is in fact comparable to the values found in Figure 4.6 (around $-14 \text{ kJ}\cdot\text{mol}^{-1}$). These results highlight the significance of anion- π interactions for the heavier anions. The fluoride potential profile is purely repulsive. This may be due to a dominance of the effect of the solvent. In Figure 4.8, one has removed the dispersion contributions to the energy. In this case, all the potentials are shown to be repulsive. The effect is much stronger in the iodide case, with a drastic change of the curve particularly in the region of the minimum at 3.9 \AA . It is also possible to note that by removing dispersion effects the profile of the potential curves for chloride, bromide and iodide are almost identical. The major difference is in the onset of the strong repulsion curve, which happens slightly earlier for heavier anions. This is maybe related with the anion size.

One repeated almost the same set of calculations for the guanidinium group, taking the central carbon atom as a reference and approaching the anions along the axis. In

these calculations one used fluoride, chloride and iodide. The results are presented in Figure 4.9.

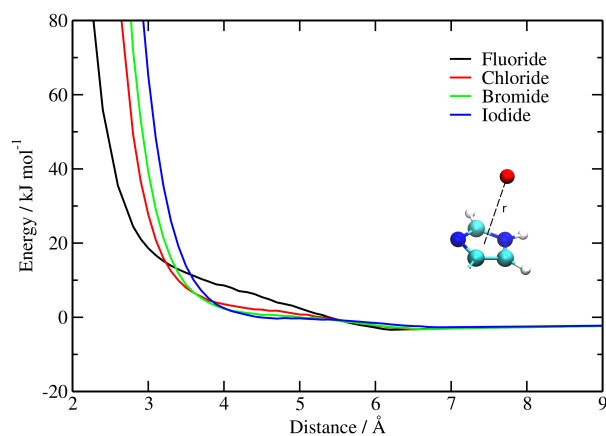


Figure 4.8: Potential energy curve (in kJ·mol⁻¹) with COSMO-LMP2 without dispersion interactions for the neutral histidine system in combination with the anions F⁻, Cl⁻, Br⁻ and I⁻, relative to the distance r along the ring normal axis.

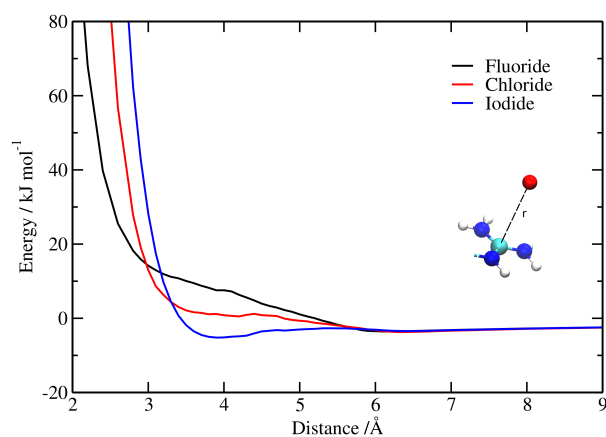


Figure 4.9: Potential energy curve (in kJ·mol⁻¹) with COSMO-LMP2 for arginine system in combination with the anions F⁻, Cl⁻ and I⁻, relative to the distance r along the guanidinium normal axis.

The obtained energy profiles show the same behaviour as in the case of the histidine ring. Fluoride and chloride present a pure repulsive behaviour towards the above plane of the guanidinium plane. On the other hand, iodide presents a minimum around 3.9 Å with a depth of $-5.9 \text{ kJ}\cdot\text{mol}^{-1}$. As observed for the histidine ring,¹²³ upon the removing dispersion effects (see Figure 4.10), the profiles depict a pure repulsive behaviour. This once more empathizes the role of dispersion effects in the interactions of anions with π -systems.

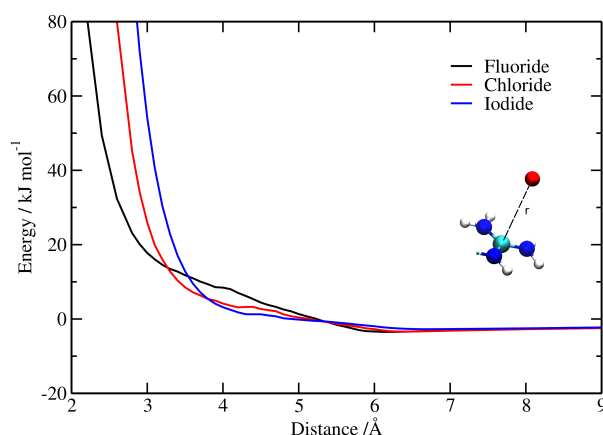


Figure 4.10: Potential energy curve (in $\text{kJ}\cdot\text{mol}^{-1}$) with COSMO-LMP2 without dispersion interactions for arginine system in combination with the anions F^- , Cl^- and I^- , relative to the distance r along the guanidinium normal axis.

Final Remarks

In this Chapter, the PES surface of different amino acids models were calculated with the LMP2 method including solvation effects through the COSMO model. The obtained profiles show the increase of affinity from iodide to fluoride towards the charged groups of the different amino acid models. PES for fluoride present sharp and deep minima next to charged groups, which suggests higher residence times in these more hydrophilic areas. Bromide and iodide, on the other hand, present more shallow minima which spread to

hydrophobic regions of the amino acid. This suggests lower residence times next to charged groups, with higher mobility to visit more hydrophobic regions in peptides.

The previous statement is also supported when analysing a backbone model and the anion- π interactions. The PES for iodide and histidine shows that 50 % of the depth in the sharpest minimum of the iodide with the imidazolium group is due to dispersion effects. For fluoride, however, dispersion forces contribute with only 5 %. The competition of large anions with water for the amino acid charged groups should also promote their interaction with hydrophobic groups of the former. For small anions like fluoride, this competition is not expected, and the stronger electrostatic interactions coupled to a weak contribution from dispersion should prohibit significant interactions with hydrophobic groups. The high affinity of the small anions for polar amino acids (or more hydrophilic parts), allows the backbone of the protein to be able to fold and to aggregate with other proteins leading to the protein salting-out, Figure 4.11-a). Contrary to the latter, large anions are able to interact more hydrophobic parts of the amino acid. This allows them to visit regions very next to the backbone, Figure 4.11-b), constraining the latter. As a consequence, they prevent the backbone folding and therefore contribute to the protein solubilization.

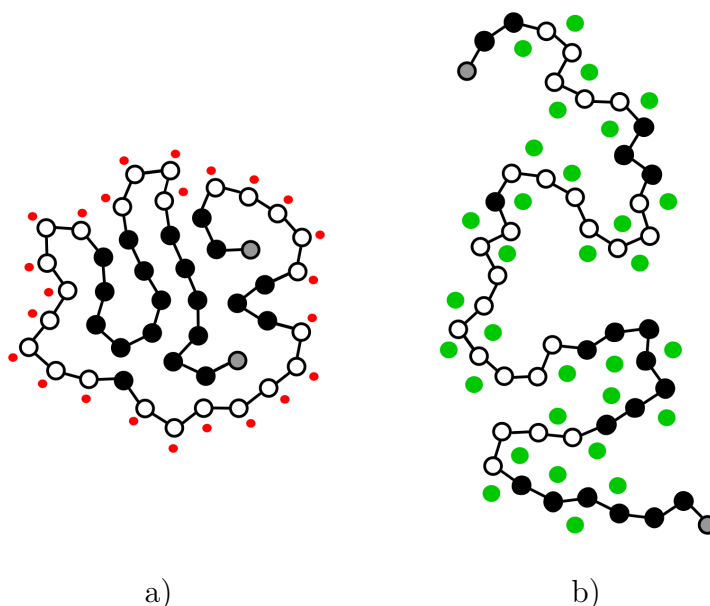


Figure 4.11: Schematics for a possible model of Hofmeister-type effects, where small anions (in red) and big anions (in green) interact with polar (white circles) and apolar (black circles) amino acids of a protein and influence its folding.

This work provides a solid picture of the molecular origin of specific anion effects and their ordering in the Hofmeister series. On one end, the results support the scenario where large polarizable anions can interact with hidden hydrophobic groups in the protein. This promotes protein destabilization and their solubilization. On the opposite end, the inability of small anions to visit hydrophobic regions, due to a high affinity for charged groups in the protein, should promote the dehydration of the latter, stabilizing it and triggering slating-out.

5 β -Peptides conformational studies

5.1 Overview

β -amino acids are non-natural occurring organic molecules which, just as their α -counterparts can build oligomeric structures. The difference lies in an extra methylene group between the C-terminus and the N-terminus of the amino acids. A general scheme for both types α - and β -amino acids is present in Figure 5.1. This carbon can be designated as carbon beta (C^β) or carbon three (C^3). The latter results from counting the carbon from the carboxylic acid (C^1) in direction of the amine group. This additional spacer between the peptidic bonds (in a peptide) gives β -peptides additional degrees of freedom and generates a high structural diversity.¹²⁴ Besides enantiomeric isomers, the possible substitution of the hydrogens of the C^β or of C^α (β^3 - and β^2 -homoamino acids) results in structural isomers.^{124–126} Cyclic side chains that share the C^α and C^β are also possible.^{124–129} β -peptides are known to adopt several conformations that can be characterized by internal dihedral angles, ϕ , θ and ψ defined as $\phi = C-N-C^\beta-C^\alpha$, $\theta = N-C^\beta-C^\alpha-C$ and $\psi = C^\beta-C^\alpha-C-N$.¹²⁵ Their secondary structures can be generally defined as helical and sheet like conformations.^{125,127} Five types of helical conformations have been identified, namely the 8-, 10-, 12-, 14- and 10/12-helix.^{125,127,130} The cardinal numbers indicate the number of atoms between hydrogen bond membered rings formed in the backbone as shown in Figure 5.2.

In the 14-helix the pattern is originated by the interaction between the amide hydrogen (N-H) from the i th residue and the oxygen from the amides carbonyl group of the $(i+2)$ th residue.^{99,125,127} The discovery of the 14-helix can be attributed to Gellman¹³¹ and to Seebach¹³² and their co-workers. The first showed that *trans*-2-aminocyclohexanecarboxylic acid (*trans*-2-ACHC) adopts a 14-helix conformation in methanol as well as in the solid state. Seebach suggested that an hexapeptide with β^3 -hVal, β^3 -Ala and β^3 -hLeu formed a 14-helix in pyridine where the side chains were atop each other. Due to the backbone

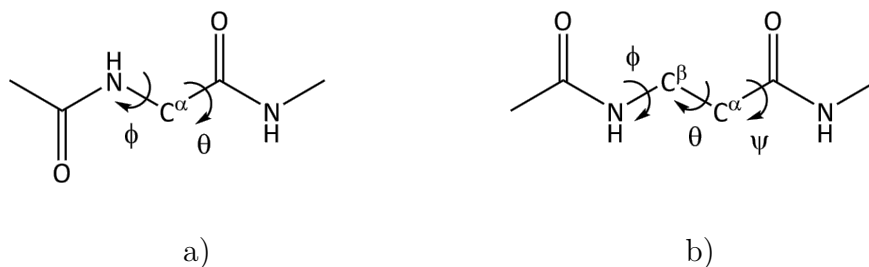


Figure 5.1: Schematic representation of the backbone of a dipeptide of a) α -amino acids and b) β -amino acids and definition of their internal dihedral angles. The latter are defined as $\phi = \text{C-N-C}^\beta\text{-C}^\alpha$, $\theta = \text{N-C}^\beta\text{-C}^\alpha\text{-C}$ and $\psi = \text{C}^\beta\text{-C}^\alpha\text{-C-N}$.

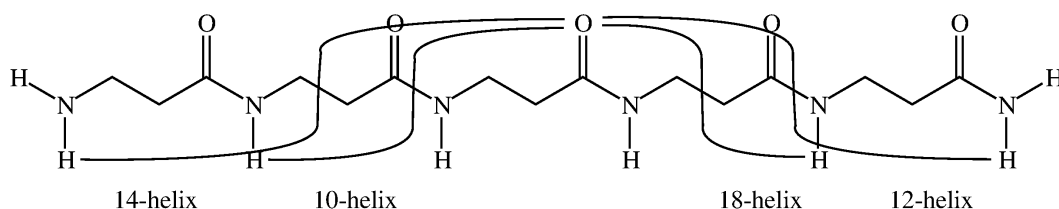


Figure 5.2: Schematic representation of some of the hydrogen bond pattern adopted by β -peptides secondary structures.

constraints, β -peptides are known to form a stable 14-helix with four residues.¹²⁷

In the case of the 12-helix, the hydrogen bonds are between the oxygen from the amides carbonyl group of the i th residue and the amide hydrogen (N-H) from the $(i+3)$ th and contrary to other structures it requires a θ dihedral angle of 90 degrees.^{99,125,127} This conformation presents the particularity that the internal hydrogen bond orientation and macrodipole are similar to the ones from α -helix from α -peptides.¹³³

The 10/12-helix is somewhat different, as it presents intersperse 10- and 12-membered ring hydrogen bonds. Contrary to the other helical structures in which all the carbonyl groups point in the same direction and the amide hydrogens in the opposite direction, in the 10/12-helix the carbonyl groups point in two alternate directions.^{99,125,127} This conformation is composed by residues that are alternated β^2 - and β^3 -amino acids.

In sheet like conformations are structures like hairpin,¹³⁴ C6-ribbon⁹⁹ or 6-strand.¹²⁵ The latter are stabilized by 6 membered ring hydrogen bonds. An example of this type of conformation is reported by Martinek and co-workers relative to *cis*-2-aminocyclopentane carboxylic acid (*cis*-ACPC) oligomers.¹³⁵ The authors observed experimentally by NMR cross-relaxation data and circular dichroism in methanol and by molecular dynamic sim-

ulations using the MM3 force field that this type of oligomers adopted a C6-strand like conformation.¹²⁵ This supports the suggestion that $\beta^{2,3}$ -amino acids with a *like* (*syn*) conformation tend to fold into sheet like conformations.¹²⁷

Contrary to *cis*-ACPC, the *trans*-ACPC, in which the side chain is in an *unlike* (*anti*) conformation, it was found to adopt a very stable 12-helix conformation in organic solvents, proved experimentally by circular dichroism and X-ray diffraction.¹³⁶ This suggests that through the design of the side chains conformations one can have some control on the preferential conformations.¹²⁹ The reason of the stability of the C6-strand is believed to be related with the *gauche* conformation of the θ dihedral angle. This is stabilized by electrostatic interactions between the negatively charged nitrogen and the positively charged carbon from the backbone's amide.⁹⁹ This sheet like conformation has also been studied in a combination of side by side strands structures.^{137–139}

The strands can be classified as polar or apolar. In the former all the carbonyl groups from the backbone point all in the same same direction, while in of apolar strands, the carbonyl are facing alternate directions.¹³⁷ Theoretical calculations with geometry optimization at the B3LYP/6-31G* level followed by energy refinement with a higher level of theory have been applied to the conformation stability of this systems, and shown to reproduce experimental results.^{137–139}

Martinek *et al.* states that hydrogen bonds that stabilize the β -peptides secondary structures can have parallel or antiparallel orientations in relation to the direction of the peptide chain, this is, from the N-terminus to the C-terminus.¹²⁵ This orientation is connected to the pattern of the hydrogen bond membered ring. In the parallel orientation, the amide hydrogens (hydrogen bond donor) are pointing in the direction of the C-terminus, while the carbonyl oxygens (hydrogen bond acceptor) point into the N-terminus. This is observed for the 10- and 14-helix and C6-strand. In the case of antiparallel, the hydrogen bond acceptor and donor point in the opposite direction. This is observed for the 8- and 12-helix.

In addition to the hydrogen bond pattern to distinguish the β -peptides conformation, one can also characterize them in terms of helical twist. Seebach and co-workers use in their adopted nomenclature (M) and (P) to differentiate left from right handed conformations.¹²⁸ Besides the nomenclature used so far in this manuscript to describe β -peptides secondary structures other variants can also appear in the literature.¹²⁷ One helical handedness can be obtained just by carrying out a mirror operation. This would correspond to multiplying their dihedral angles by -1 on a conformation with oppo-

site helical handedness and by inverting the configuration of the C^α and C^β .¹²⁵ Several nomenclatures have been used for the secondary structures of β -amino acids.¹²⁷

In terms of the stability of unsubstituted β -peptides conformers, the 10- and 12-helix are more stable than the 14-helix in gas phase. Nevertheless, in solution the 14-helix is most stable than the other conformations. This effect was observed for helical structures with more than four residues. The destabilization of the previous conformations over the 14-helix is believed to be related to their reduced dipole moment.⁹⁹ The effect of different side chains has been investigated by Lin *et al.*¹⁴⁰ using molecular dynamic simulation using GROMOS force field on β -heptapeptides. In the latter the fifth residue was changed into different branched (like β^3 -hVal) and non-branched (like β^3 -hAla) side chains.¹⁴⁰ The authors come to the conclusion that the dominant conformation was the 14-helix. It was found that there was a very small stabilization ($2 \text{ kJ}\cdot\text{mol}^{-1}$) due to the presence of branching side chains on the C^γ (carbon of the side chain that is directly connected to the backbone), but no effect was found due to the variation of the length of the side chain or its polarity.¹⁴⁰

The work presented in this chapter is dedicated to the study of β -peptides stability in several secondary conformations. The latter are composed by a hexamer of β -amino acids in a β -sheet like, 12- and 14-helix structures. For this effect a set of four amino acids were selected, with cyclic and aliphatic side chains. The conformational studies were carried out using quantum mechanical methods in gas phase and in solution. Solvent effects were simulated using an implicit solvation model.

5.2 Computational Details

For the β -peptides conformational studies, one built β -amino acid hexamers. The latter are mainly divided in two groups. One where the amino acids present a cyclic side chain, while the second is constituted by aliphatic side chain. The structures of the β -amino acids used in this work are present in Figure 5.3.

For the first set, the model structures of *cis*-2-aminocyclopentanecarboxylic acid (*cis*-ACPC) and *cis*-3-aminotetrahydrofuran-2-carboxylic acid (*cis*-FAA) hexamers were built on the basis of experimental structural studies¹⁴¹ with a β -sheet (C6 strand like structure) and a 14-helix left handed conformations.

The second test set is composed of β^3 -hVal and of β^3 -hLeu conformers built in a

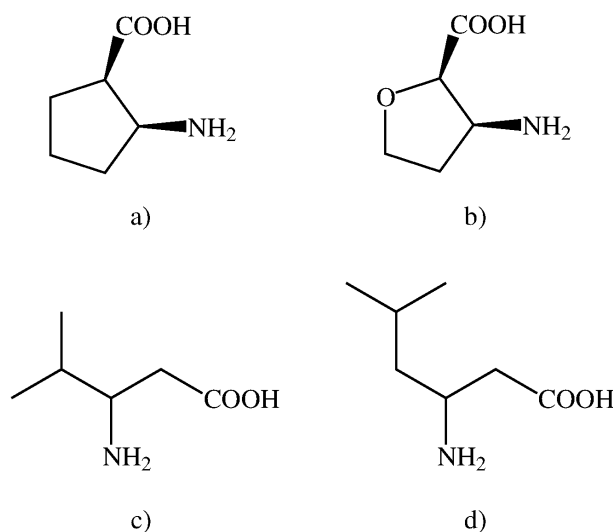


Figure 5.3: Structure of the β -amino acids used in the present study. The former are composed of: a) *cis*-2-aminocyclopentanecarboxylic acid (*cis*-ACPC), b) *cis*-3-aminotetrahydrofuran-2-carboxylic acid (*cis*-FAA), c) β^3 -hValine and d) β^3 -hLeucine.

14-helix and 12-helix, both in a right (P) and left (M) handedness. For each of the four structures one built all the possible configurations of the side chain according to the conditions referred by Seebach *et al.*¹²⁸ Here for a 14-helix conformation only the axial positions of the side chain relative to the helix are possible. Taking this into account in the case of β^3 -hVal only 3 configurations of this side chain are possible. In the case of β^3 -hLeu this number increases to 9 for each helical structure, as shown in Figure 5.4.

The oligomers were capped in the N-terminus with azide and in the C-terminus with methyl ester. In order to generate suitable starting structures, the initial handmade geometries were optimized with RM1 semiempirical method as implemented in MOPAC quantum chemistry program package.¹⁴²

The obtained structures for the peptides with cyclic side chains were then optimized at the B3LYP/6-31G* level of theory in gas phase and in solution.⁵¹ The DFT optimized structures were also used in estimating starting geometries for the *cis*-FAA hexamers. This was accomplished by substituting a CH₂ in the cyclopentane rings by an oxygen, according to Figure 5.3. This was followed by geometry optimizations at the same level of theory.

For the peptides with an aliphatic side chain, the structures were optimized at the B3LYP/6-31G* level of theory.⁵¹ A set of geometries were then selected in a range of 50 kJ mol⁻¹ in relation to the most stable conformation and optimized again with

B3LYP/6-31G* including dispersion corrections,⁵³ (B3LYP-D3), for gas phase and in solution.

LMP2 single points with density fitting approximations were carried out in each structure.⁴³ Single points with B3LYP with and without dispersion corrections⁵³ were also carried out. Energy single points were both in gas phase and in solution. The solvent effects were introduced both in geometry optimizations and energy single points through the use of the COSMO model.⁶ The former include trifluoroethanol (TFE) and water (W) with dielectric constants of $\epsilon=26.14$ and 80.10 respectively.

For the peptides with cyclic side chains, the basis set used was the Dunning cc-pVTZ basis¹⁴³ with diffuse functions added to non-hydrogen atoms.⁹⁰ This will be hereafter referred to as aug'-cc-pVTZ. While for the peptides with aliphatic side chains were computed with with a cc-pVTZ basis set.⁹⁰

In the local calculations, Pipek-Mezey orbitals³⁴ were used in combination with Natural Population Analysis (NPA) domain selection criteria, with $T_{\text{NPA}}=0.03$.³⁷ LMP2 calculations were carried out with a development version of Molpro 2010.2.³⁸ Single points using density functional theory were performed with the Orca quantum chemistry software package.¹⁴⁴

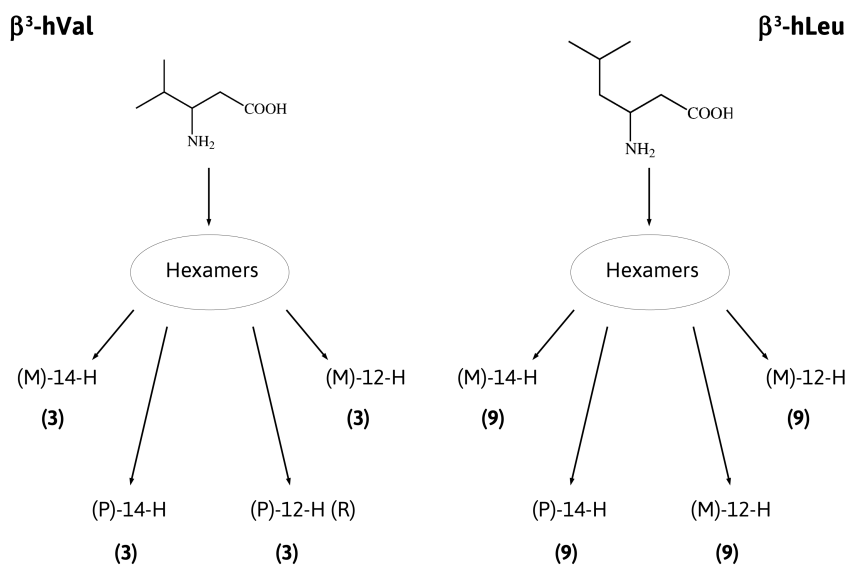


Figure 5.4: Schematic representation of the study of the stability of β^3 -hVal and of β^3 -hLeu hexamers in a right (P) and left (M) handed conformations. The cardinal number in parenthesis indicates the number of structures built.

5.3 Results and Discussion

5.3.1 β -peptides with cyclic side chains

The focus of this work is to study the stability of *cis*-ACPC and *cis*-FAA in β -sheet and 14-helix conformations. The initial structures of the *cis*-FAA were created from the optimized geometries of the *cis*-ACPC with the above mentioned changes. This procedure was applied in order to maximize the similarities in the secondary structures of the oligomers of the two β -amino acids. This allows a comparison between the two hexapeptides on a relative equal footing, while avoiding an extensive minima search. This would be prohibitive using quantum mechanics due to the system size, even if semiempirical methods were to be used. Alternatively one could opt for the use of molecular mechanics. Although recently efforts have been made in obtaining reliable force field parameters for the description of β -peptides, such as in the case of *cis*-ACPC,¹⁴⁵ there is no consistent set available for both monomers considered in this study.

Both the β -sheet and 14-helix structures for the *cis*-ACPC and *cis*-FAA hexamers were optimized in gas phase and in solution. In solution two different polarity environments were used. They were included using a continuum description of trifluoroethanol and water. In Figure 5.5 one presents the optimized geometries of β -sheet and 14-helix conformations of the two β -amino acids in study obtained for gas phase. Allylic hydrogens were removed for a better visualization.

One will start by discussing the optimized 14-helix structures. A comparison between the *cis*-ACPC and *cis*-FAA shows that the differences are very subtle. A first observation would be that the *cis*-ACPC hexamer presents a hydrogen bond between the methyl ester cap carbonyl group and the amide hydrogen from the ($i+2$)th residue. Such was observed for all three simulated environments. This hydrogen bond presents the same characteristic of the hydrogen pattern observed in a 14-helix. This formation was found to be absent in the case of the *cis*-FAA conformer. This is due to the fact that the ester group in *cis*-ACPC presents an eclipsed conformation, instead of a gauche conformation as it is observed in the *cis*-FAA oligomer. It is difficult to judge whether this is an effect directly linked to the substitution in the ring or an artifact from the geometry optimization. A possible way to overcome this doubt would be to perform an extensive study of the conformational landscape, using several structures for each conformer. It is important to note that such changes in the terminal groups can easily happen since they

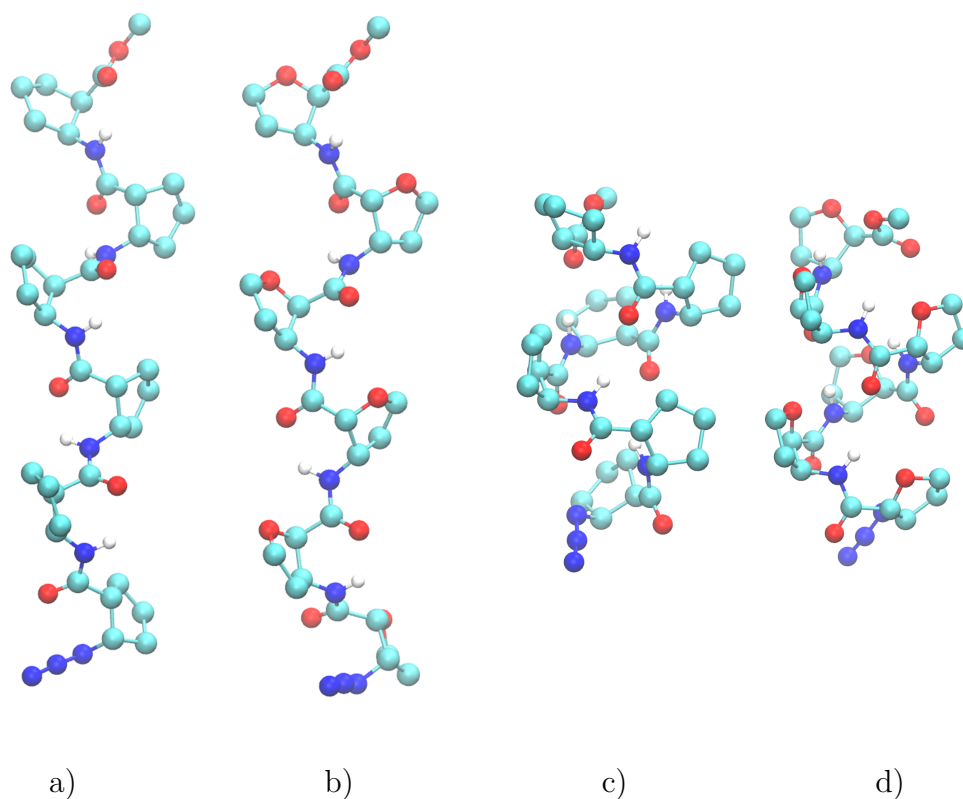


Figure 5.5: Structures of the most stable conformations for the *cis*-ACPC hexamer in a a) β -sheet and c) 14-helix and for the *cis*-FAA hexamer in b) β -sheet and d) 14-helix . The geometries were optimized with B3LYP-D3/6-31G* in trifluoroethanol (TFE). The solvent effects were introduced by using the COSMO model. Allylic hydrogens were removed for clarity.

exhibit a higher flexibility. Nonetheless, they should have a small effect in the relative conformational energies.

An average value for the hydrogen bond distances and angles is presented in Table 5.1. It can be noticed that for the gas phase the hydrogen bond length of *cis*-FAA is 0.05 Å smaller than *cis*-ACPC. The angles are basically the same. When one goes to solution more significant differences are observed. *cis*-FAA presents hydrogen bond 0.16 Å smaller than *cis*-ACPC and smaller angles by about 8 degrees. This observation suggest that is only in solution that there are significant differences between the hydrogen bonds of *cis*-ACPC and *cis*-FAA 14-helices structures. Smaller hydrogen bond lengths for *cis*-FAA suggest that it presents a tighter packing than *cis*-ACPC. If one compares the 14-helix *cis*-ACPC conformers, one notices that the major differences occur between the gas

phase and solution geometries. No significant differences are observed for the hydrogen bond lengths and angles between the two simulated solvents. The previous evidences are also observed for the *cis*-FAA.

Table 5.1: Average hydrogen-bonds lengths (in Å) and angles (in degrees) present in *cis*-ACPC and *cis*-FAA 14-helix conformations. For the optimized structures in the gas phase (GP), trifluoroethanol(TFE) and water (W) with B3LYP/6-31G*.

	<i>cis</i> -ACPC		<i>cis</i> -FAA	
	Distance	Angle	Distance	Angle
GP	2.834	151.6	2.697	151.6
TFE	2.373	161.5	2.205	153.2
W	2.363	161.9	2.218	153.8

The side chain in the *cis*-FAA and *cis*-ACPC in a 14-helix conformation exhibits a strong influence on the tightness of the helix loop. Comparing the distances between the amide hydrogen from the backbone and the ring oxygen in *cis*-FAA and the carbon in *cis*-ACPC (in the equivalent position of the oxygen in the former) the ones in the former case are found to be much shorter (Figure 5.6). In the *cis*-FAA 14-helix the hydrogen and the oxygen are 2.1-2.1 Å far apart. In the case of *cis*-ACPC the distance between the hydrogen and the carbon is found to be 2.5-2.6 Å. The reason for this is obvious. By replacing a CH₂ group with an oxygen in the cyclic side chain, the steric effects are reduced and also introduces a favourable electrostatic interaction between the latter and the hydrogen. This favourable interaction contributes to a more compressed helical structure which is also reflected in shorter hydrogen bonds in the 14-helix pattern.

One also calculated the solvent accessible surface area (SASA) for these conformers in the three media. Calculating the difference between the SASA of *cis*-FAA and *cis*-ACPC shows that in gas phase this is 6 % smaller for the former . The same trend is observed in solution, being 8 % smaller for both simulated solvents. This suggests that the *cis*-FAA 14-helix exhibits a more compact conformation.

The previous effect should also be reflected in the backbone properties such like the dihedral angles. This is presented in Table 5.2. The ψ angles are usually taken as the

determining parameters for the backbone conformation.¹⁴⁶ It can be observed that for the gas phase the *cis*-FAA 14-helix presents a smaller ψ angle. This supports the idea of a tighter packing in the *cis*-FAA in good agreement with the SASA values.

The effect of the medium on the optimized geometries for the 14-helix of each β -amino acids was also assessed through the root mean square deviation (RMSD), Table 5.5. The superposition of the hexamer structures for *cis*-ACPC is presented in Figure 5.7 and for *cis*-FAA in Figure 5.8. The allylic hydrogens were removed for better visualization. For the *cis*-FAA 14-helix a significant difference in the RMSD is only observed between the gas phase and solution geometries, with variations of 0.45 Å for trifluoroethanol and 0.42 Å for water. Comparing the two simulated solvents this difference is more subtle, 0.12 Å. In the case of *cis*-ACPC 14-helix the same trend is observed. This reveals that the absence or presence of solvent has some impact on the geometry, while there is only very small effect between the two solvents.

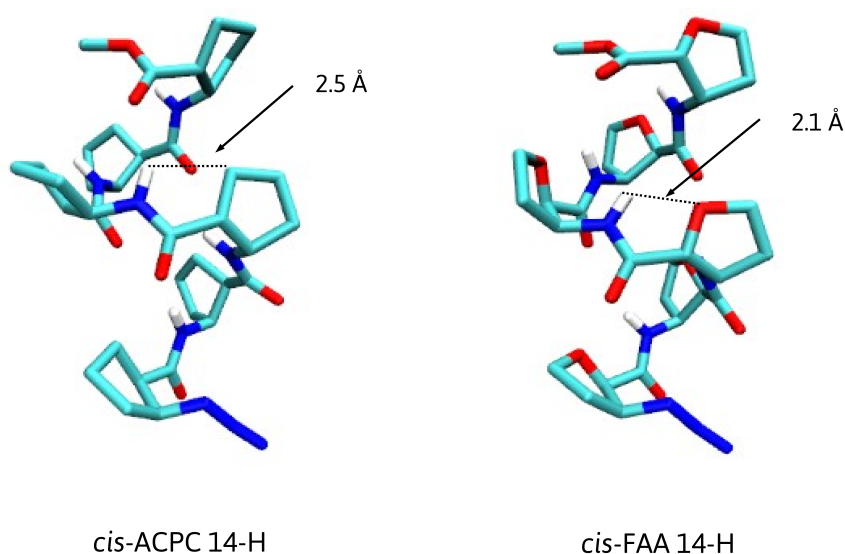


Figure 5.6: Schematic representation of favourable intrinsic interactions that stabilizes the *cis*-FAA 14-helix that is absent in the case of the *cis*-ACPC.

Now one will compare the β -sheet conformers for *cis*-FAA and *cis*-ACPC. In this type of conformation the hydrogen bond pattern results from the interaction of the amide hydrogen and the oxygen of the carbonyl group that belong to the same amino acid. In this case the donor and acceptor sites are separated by 6 atoms. Their average length and angles are presented in Table 5.3. One can notice that, as in the case of the 14-helix, the

Table 5.2: Average backbone dihedral angles (in degrees) for 14-helix conformation of *cis*-ACPC and *cis*-FAA hexamers. Values for the geometries optimized in gas phase (GP), trifluoroethanol (TFE) and water (W). The torsional degrees of freedom are defined as follows: $\phi = \text{C-N-C}^\beta\text{-C}^\alpha$, $\theta = \text{N-C}^\beta\text{-C}^\alpha\text{-C}$ and $\psi = \text{C}^\beta\text{-C}^\alpha\text{-C-N}$.

	<i>cis</i> -ACPC			<i>cis</i> -FAA		
	ϕ	θ	ψ	ϕ	θ	ψ
GP	-138.7	46.1	-124.1	-145.9	38.7	-114.6
TFE	-130.3	43.3	-128.4	-142.6	34.4	-110.8
W	-122.1	44.0	-128.6	-140.7	35.0	-112.4

cis-FAA presents shorter hydrogen bonds than *cis*-ACPC. When going from gas phase to solution this difference increases approximately by 0.07 and 0.19 Å in TFE and W, respectively. Comparing the angles one can notice that they are quite comparable for gas phase while in solution *cis*-FAA presents angles up to 12 degrees larger than *cis*-ACPC in water. The smaller hydrogen bond lengths registered seem to be related to a more tight conformation as it was observed in the case of the 14-helix.

Table 5.3: Average hydrogen-bonds lengths (in Å) and angles (in degrees) present in *cis*-ACPC and *cis*-FAA β -sheet conformations. For the optimized structures in gas phase (GP), trifluoroethanol(TFE) and water (W) with B3LYP/6-31G*.

	<i>cis</i> -ACPC		<i>cis</i> -FAA	
	Distance	Angle	Distance	Angle
GP	2.155	120.4	2.089	124.5
TFE	2.216	118.5	2.022	129.0
W	2.209	116.9	2.025	129.0

In order to have an idea of the structure packing one can again take a look into the dihedral angles. The average dihedral angles for the *cis*-FAA and *cis*-ACPC in a β -sheet

conformation are presented in Table 5.4. In general, the same trend is observed in the three simulated environments. The θ angles for *cis*-FAA and *cis*-ACPC are quite comparable. In the case of ψ , smaller values were obtained for *cis*-FAA indicating that this presents a more compact packing. This idea is supported by the obtained SASA values. The SASA of *cis*-FAA in gas phase is 3 % smaller than *cis*-ACPC. In solution this difference reaches 6 % for both solvents.

Table 5.4: Average backbone dihedral angles (in degrees) for β -sheet conformation of *cis*-ACPC and *cis*-FAA hexamers. Values for the geometries optimized in gas phase (GP), trifluoroethanol (TFE) and water (W). The torsional degrees of freedom are defined as follows: $\phi = \text{C-N-C}^\beta\text{-C}^\alpha$, $\theta = \text{N-C}^\beta\text{-C}^\alpha\text{-C}$ and $\psi = \text{C}^\beta\text{-C}^\alpha\text{-C-N}$.

	<i>cis</i> -ACPC			<i>cis</i> -FAA		
	ϕ	θ	ψ	ϕ	θ	ψ
GP	-111.0	-41.9	159.1	-119.8	-41.4	146.7
TFE	-110.8	-41.8	149.9	-132.7	-43.7	140.0
W	-112.0	-41.9	160.9	-132.9	-43.9	140.5

The effect of the medium on the optimized geometries for the β -sheet of the hexamers was also assessed through the root mean square deviation (RMSD). The values are given in Table 5.5. The superposition of the hexamer structures for *cis*-FAA is presented in Figure 5.8 and for *cis*-ACPC in Figure 5.7. The allylic hydrogens were removed for clarity. For the *cis*-FAA β -sheet a significant difference is observed between the gas phase and solution geometries, with variations of 0.76 Å for trifluoroethanol and 0.75 Å for water. Comparing the two simulated solvents this difference is more subtle, 0.03 Å. In the case of *cis*-ACPC β -sheet the same trend is observed. This reveals that the solvent has some impact on the geometry, while there is only a very small difference between the two solvents.

Table 5.5: RMSD for *cis*-ACPC and *cis*-FAA in the β -sheet and 14-helix optimized geometries (in Å) obtained in gas phase (GP), trifluoroethanol(TFE) and water (W).

	<i>cis</i> -ACPC		<i>cis</i> -FAA	
	β -sheet	14-H	β -sheet	14-H
GP <i>vs.</i> TFE	0.39	0.32	0.76	0.45
GP <i>vs.</i> W	0.40	0.39	0.75	0.42
TFE <i>vs.</i> W	0.05	0.03	0.03	0.12

The relative energies for *cis*-ACPC and *cis*-FAA hexapeptides are displayed in Table 5.6. The values are given as the energy difference between the 14-helix and the β -sheet conformations, $E(14\text{-helix}) - E(\beta\text{-sheet})$. Positive values in the table indicate a larger stabilization in the β -sheet case, whereas negative values indicate that the 14-helix conformation is more stable. The results obtained also include COSMO corrections. In the case of trifluoroethanol and water, the COSMO model was used both in the optimization procedure and in the energy refinement.

In the case of *cis*-ACPC, B3LYP and LMP2 favour a β -sheet conformation in the three simulated environments. The LMP2 stabilization energies vary from 6.0 kcal·mol⁻¹ in gas phase and 3.3 kcal·mol⁻¹ in trifluoroethanol. The relative energies are almost unchanged in going from trifluoroethanol to water. B3LYP predicts a stabilization of the β -sheet in a range of 17.6 kcal·mol⁻¹ in gas phase and 16.7 kcal·mol⁻¹ in water. In the two simulated solvents the β -sheets are stabilized to about the same extent. These observations are in good agreement with experimental^{135,141} and theoretical results.^{138,141}

The considerable differences between LMP2 and DFT can be attributed to the well known deficiency of the latter to account for dispersion effects. In order to overcome this disadvantage of DFT when compared to wavefunction methods, the energy was refined by carrying out B3LYP single points with dispersion corrections (B3LYP-D3), as proposed by Grimme and co-workers.⁵³ The obtained relative energies do not seem consensual. In gas phase the most favourable conformation is the β -sheet. When one introduces solvent effects the most stable conformation is the 14-helix.

The LMP2 method with the same basis set as used in this work has been applied to the

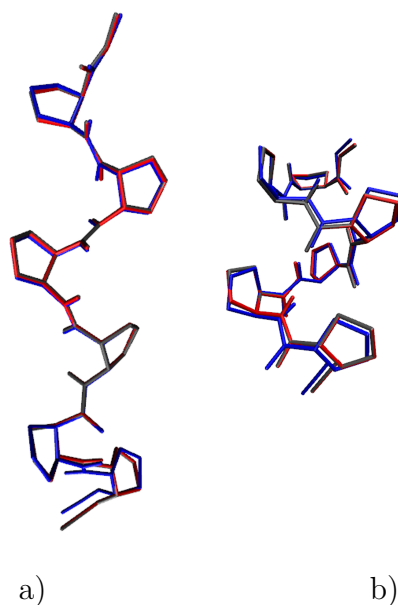


Figure 5.7: Superimposed optimized structures for the *cis*-ACPC β -sheet (left) and *cis*-ACPC 14-helix (right) under different solvent descriptions. The structures are color coded distinguishing gas phase (blue) from trifluoroethanol (red), and water (grey)

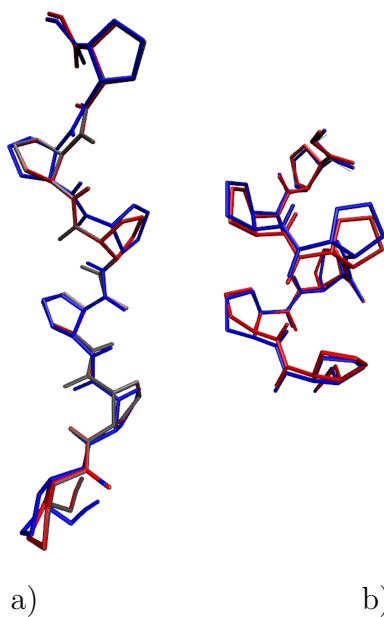


Figure 5.8: Superimposed optimized structures for the *cis*-FAA β -sheet (left) and *cis*-FAA 14-helix (right) under different solvent descriptions. The structures are color coded distinguishing gas phase (blue) from trifluoroethanol (red), and water (grey).

conformational study of α -alanine tetra-, hexa- and octamers.¹⁰¹ The energy differences of the previous method compared to the basis set limit were shown to be less than 1 kcal·mol⁻¹. Taking this into account it may be expected that the results at the LMP2 level of theory can be considered as a reference.

Table 5.6: Relative energy differences (in kcal·mol⁻¹) between the β -sheet and the 14-helix conformers computed at the DF-LMP2, B3LYP and B3LYP-D3 methods with the aug'-cc-pVTZ basis set [E(14-helix)-E(β -sheet)]. The structures were optimized with B3LYP/6-31G*. The COSMO model was used to approximate solvent effects for both single point and optimization runs.

	<i>cis</i> -ACPC			<i>cis</i> -FAA		
	B3LYP	B3LYP-D3	LMP2	B3LYP	B3LYP-D3	LMP2
GP	17.6	1.5	6.0	-4.3	-17.6	-16.7
TFE	16.9	-1.0	3.3	4.3	-10.0	-7.7
W	16.7	-1.4	3.4	4.4	-10.0	-8.6

In the case of *cis*-FAA, contrary to what was observed for *cis*-ACPC, B3LYP-D3 is in better agreement with LMP2 than B3LYP. They show higher stabilization of the 14-helix over the β -sheet. This conclusion is in good agreement with experimental work from Pandey *et al.*¹⁴¹ The LMP2 relative energy are in between -16.7 kcal·mol⁻¹ for gas phase and -7.7 kcal·mol⁻¹ for trifluoroethanol. For B3LYP-D3 these values are between -17.6 kcal·mol⁻¹ for gas phase and -10.0 kcal·mol⁻¹ for solution. At the B3LYP the 14-helix conformation is more favourable only in gas phase, while in solution the β -sheet conformation dominates. It is important to notice that the B3LYP-D3 energies correlate very well with the ones at the LMP2 level. The differences between the two methods vary from 0.9 kcal·mol⁻¹ for gas phase to 2.3 kcal·mol⁻¹ for triluorethanol.

These results show that LMP2 can be applied in β -peptides conformational studies, as it was in α -peptides.¹⁰¹ B3LYP and B3LYP-D3 relative energies follow same trend as LMP2 for *cis*-FAA and *cis*-ACPC with the same basis set. This shows that LMP2 is a robust method when compared with the B3LYP with and without dispersion corrections. Although B3LYP is a relatively suitable method for obtaining qualitative structures of β -peptides, one should be cautious about comparing the energetics of such system at the DFT level. This is particularly true when comparing packed and extended conformers.

The reason for the stabilization of the 14-helix over the β -sheet for *cis*-FAA like conformation may be related to its tighter packing, favouring close dispersion contacts. This is evidenced in comparing the DFT with the LMP2 results. The more compact structure of the *cis*-FAA hexapeptide should be due to a favourable interaction between the oxygen of the side chain of the residue i and the amide hydrogen from the backbone which has its origin in the $(i-1)$ th residue. Such effect is presented in Figure 5.6. This is absent in the case of *cis*-ACPC. Although this argument is in favour of the stabilization of the *cis*-FAA helical conformation, it appears to have an opposite effect for the β -sheet like conformation. This suggests that for this β -peptide this extra constrain seems to induce the helical conformation. In the case of the *cis*-ACPC the preferred conformation is the β -sheet like conformation stabilized by 6-membered hydrogen bonds. This is an apolar structure, it results from the alternate hydrogen bonds above and below the plane of the cyclic side chains. This type of conformers are mentioned as being self-stabilizing by intrinsic, according to the authors, weak hydrogen bonds.¹⁴⁷

5.3.2 β -peptides with aliphatic side chains

This section focusses on studying the stability of β^3 -hVal and β^3 -hLeu hexamers in several conformations in two different environments, gas phase and trifluoroethanol. The conformations under study were the 14-helix and the 12-helix both in a right (P) and left (M) handedness. Several nomenclatures have been used for the secondary structures of β -amino acids.¹²⁷ Here we adopt the one from Gellman¹²⁷ to describe the helical membered ring hydrogen bonds pattern, which is defined by the cardinal number, and the one from Seebach¹²⁴ to define the handedness. The most stable conformers for β^3 -hVal and β^3 -hLeu hexamers in trifluoroethanol are presented in Figure 5.9 and Figure 5.12 respectively.

Taking into account the 12-helix conformations one can notice that independently from the handedness the carbonyl oxygens from the amide in the backbone point in the direction of the methyl ester cap, while the amide hydrogens point in the direction of the N-terminus. This is in agreement with structures of β -hGly in a 12-helix conformation presented by Wu *et al.*⁹⁹ In the 14-helix, the direction of the oxygens and hydrogens of the peptidic bond is the inverse of the case of the 12-helix. With the amide oxygens pointing in the direction of the N-terminus, while the amide hydrogens point in the direction of the C-terminus. This is in agreement with the 14-helix β -hGly presented

by Wu *et al.* and for the generic 14-helices of β^3 -amino acids by Gellman *et al.*^{99,124} In the case of the β^3 -hVal 14-helices the side chains are basically almost atop of each other every 3 residues.¹²⁷ In the case of the β^3 -hLeu the 14-helices the side chains are not exactly on top of each other, presenting between them a rotation of approximately 15 degrees.¹²⁸

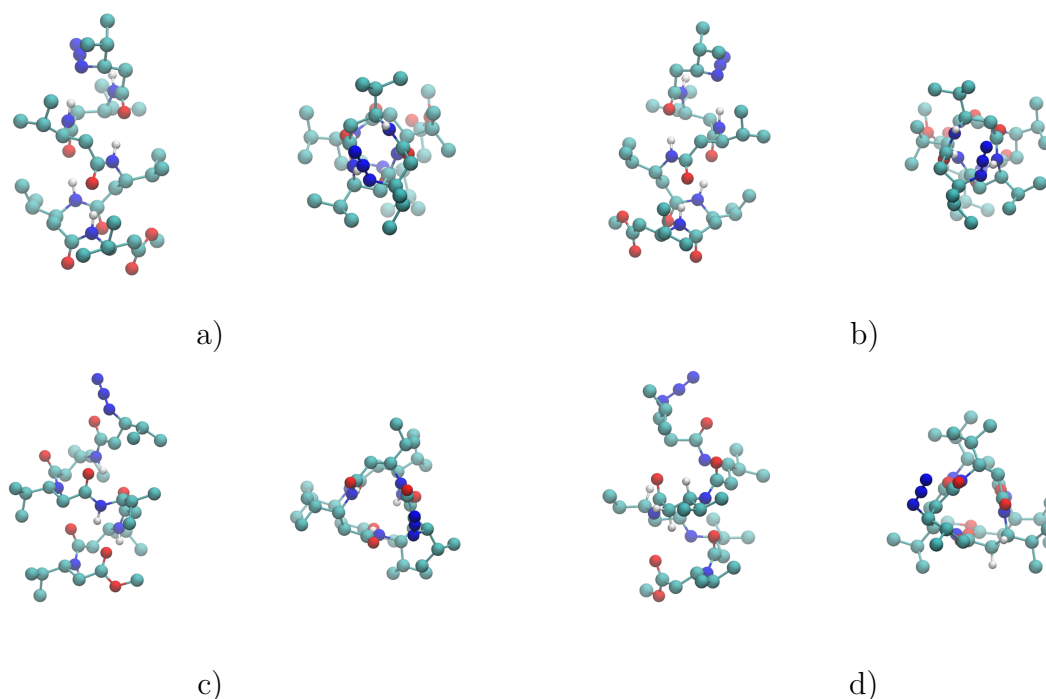


Figure 5.9: Structures of the most stable conformations for the β^3 -hVal in a 12-helix conformation a) left handed (M), b) right handed (P) and in a 14-helix conformation c) left handed (M) d) right handed (P). The geometries were optimized with B3LYP-D3/6-31G* in trifluoroethanol (TFE). The solvent effects were introduced by using the COSMO model. Aliphatic hydrogens were removed for clarity.

In order to understand the differences between the several conformations for the above mentioned amino acids, the intrinsic properties of each structure are considered. One will start by discussing the β^3 -hVal oligomers. Assessing the hydrogen bonds length, Table 5.7 one can see that for the 14-helix in gas phase they are around 2.076 and 2.139 Å for the right (P) and left (M) conformers respectively. According to Rozas, hydrogen bond lengths between 1.5-2.2 Å are considered to be strong hydrogen bonds.¹⁴⁸ Upon inclusion of solvent effects in the geometry optimizations their length is decreased by 0.17 Å. This decrease is accompanied by an increase of the hydrogen bond angle by about 2 degrees for the right handed and by about 9 degrees for the left handed structures. This seems

to indicate an accommodation of the conformers to the presence of the solvent. This is reflected by originating stronger hydrogen bonds. The same trend is observed for the 12-helix conformers. The obtained hydrogen bond lengths are in good agreement with the ones presented by Wu *et al.*, 2.21 Å for the 14-helix and 2.08 Å for the 12-helix.⁹⁹

Table 5.7: Average hydrogen-bonds lengths (in Å) and angles (in degrees) for β^3 -hVal in a 12-helix and 14-helix conformations both with a left (M) and right (P) helical folding. The structures were optimized in gas phase (GP) and in trifluoroethanol(TFE) with B3LYP-D3/6-31G*.

	GP		TFE	
	Distance	Angle	Distance	Angle
(P)-14-H	2.076	160.7	1.881	162.9
(M)-14-H	2.139	150.0	1.971	158.7
(P)-12-H	1.950	160.8	1.873	163.7
(M)-12-H	1.968	158.5	1.862	163.7

As it was shown for the cyclic β -amino acids oligomers the dihedral angles are important in order to characterize the secondary structures. The dihedral angles for β^3 -hVal in a 14-helix and 12-helix both right and left handed are presented in Table 5.8. As a first observation one can notice that one can obtain the the left handedness by multiplying the dihedral angles of the right handedness by -1. As observed by Fülöp *et al.*, a conformation with opposite helical direction can be obtained by a mirror operation and by inverting the chirality of the side chain in C $^\alpha$ and/or in C $^\beta$. This allows to discuss the dihedral angles in general terms for (M)-14-helix and (M)-12-helix since for the right handedness the conclusions are identical.

For the (M)-14-helix conformer the angle θ indicates a synclinal (gauche) conformation, this conformation allows for helix conformation.¹⁴⁹ The ϕ and ψ are in a anticlinal conformation (between 90 to 150 degrees or between -90 to -150 degrees) in good agreement with the 14-helix conformations obtained by MD using GROMOS force field⁹⁹ and for the values presented by Martinek *et al.*¹⁴⁹ for β^3 -amino acids. Comparing the dihedral angles from gas phase with the ones simulated in trifluoroethanol a slight decrease of dihedral angles is observed.

Table 5.8: Average backbone dihedral angles (in degrees) for β^3 -hVal in a 12-helix and 14-helix conformations both with a left (M) and right (P) helical folding. The structures were optimized in gas phase (GP) and in trifluoroethanol (TFE) with B3LYP-D3/6-31G*. The torsional degrees of freedom are defined as follows: $\phi = \text{C-N-C}^\beta\text{-C}^\alpha$, $\theta = \text{N-C}^\beta\text{-C}^\alpha\text{-C}$ and $\psi = \text{C}^\beta\text{-C}^\alpha\text{-C-N}$.

	GP			TFE		
	ϕ	θ	ψ	ϕ	θ	ψ
(P)-14-H	132.9	-60.0	142.5	131.9	-53.4	135.9
(M)-14-H	-149.9	56.2	-125.0	-145.3	56.1	-125.5
(P)-12-H	-102.4	91.6	-87.7	-98.1	88.3	-88.5
(M)-12-H	102.1	-89.4	84.0	97.9	-88.0	88.1

In order to have a better visualization of structural differences between the gas phase and solution geometries, one overlapped these structures. The superposition of the most stable conformers of the 14-helix in both right and left handed conformations in gas phase and in trifluoroethanol are present in Figure 5.10-b) and Figure 5.10-c) respectively.

It can be noticed that for the (M)-14-helix no significant differences are observed. This is supported by a RMSD of 0.23 Å. For the β^3 -hVal in a (M)-14-helix geometry the same conformation of the side chain revealed to be the minimum both in gas phase and in solution. In the case of the (P)-14-helix significant differences were observed, which are reflected in a RMSD of 6.05 Å. This huge discrepancy between the two structures is related to the fact that in trifluoroethanol a different conformation of the side chain was preferred. The two conformations are present in Figure 5.11. The most stable conformer in gas phase when optimized in solution is 6.2 kcal·mol⁻¹ less stable at the DF-LMP2/aug'-cc-pVTZ level than the one obtained in solution.

Keller *et al.* studied the effect of the side chain in β -peptides by molecular dynamics using the GROMOS 43A1 force field.¹⁴⁶ The peptides were constituted by β^3 -hVal and β^3 -hLeu. The authors stated that the side chains have no influence on the backbone conformation, only the amino acid sequence is relevant for the stability of a certain conformation.¹⁴⁶

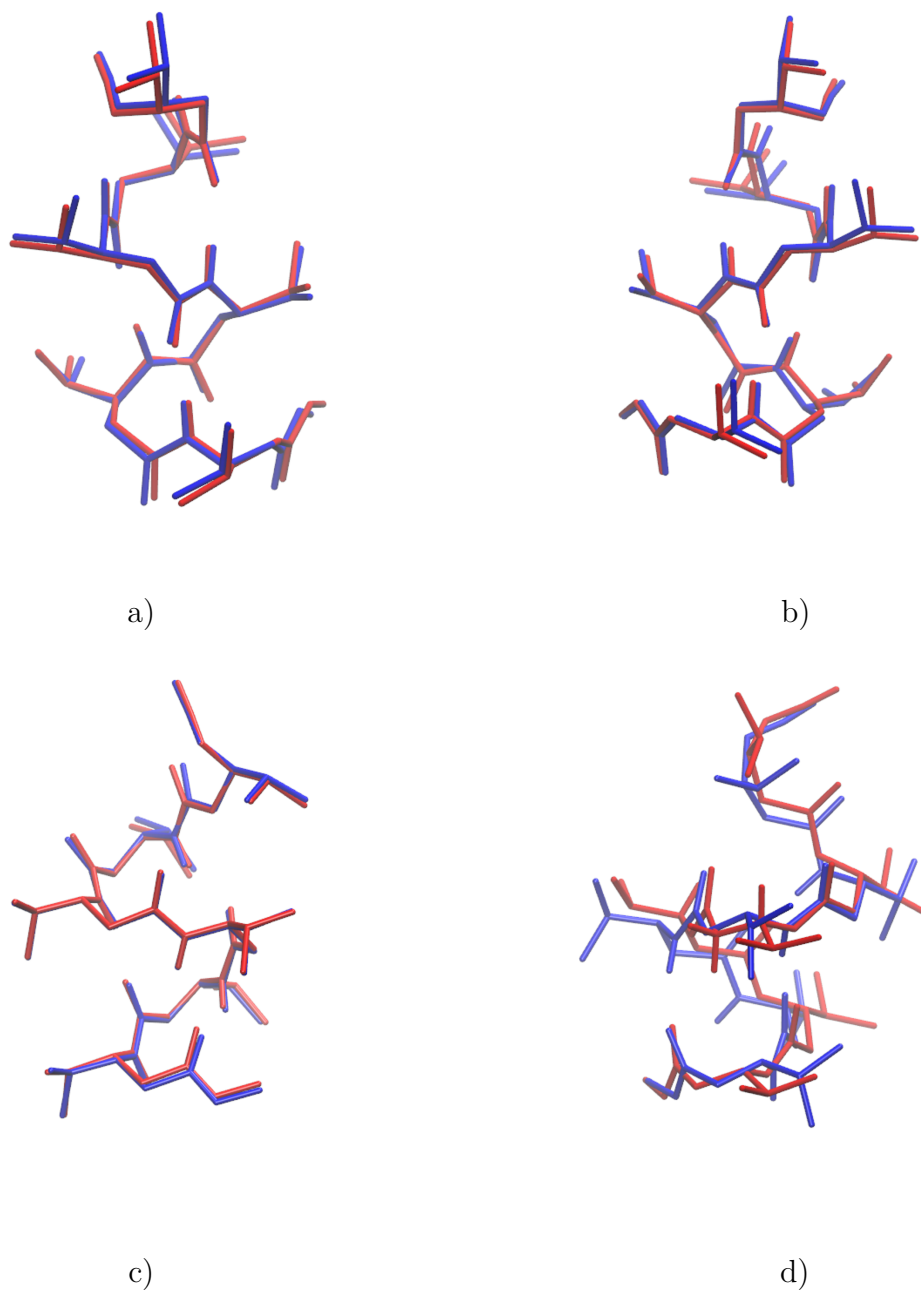


Figure 5.10: Superimposed optimized structures for the β^3 -hVal in a 12-helix conformation a) left handed (M), b) right handed (P) and in a 14-helix conformation c) left handed (M) d) right handed (P). The structures are color coded distinguishing gas phase (blue) and trifluoroethanol (red). Aliphatic hydrogens were removed for clarity.

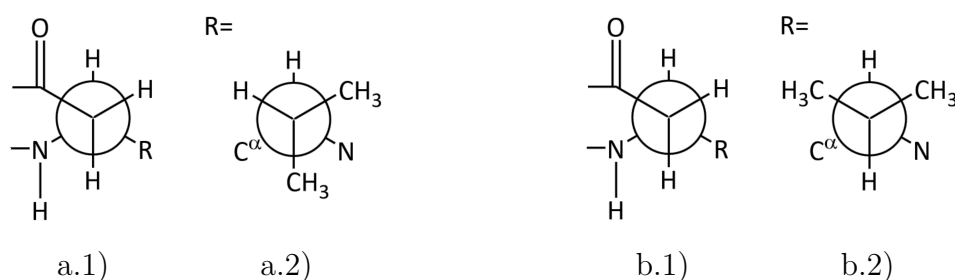


Figure 5.11: Scheme of the Newman projections of the β^3 -hVal (P)-14-helix most stable conformer in a) gas phase and in b) trifluoroethanol. 1) is the projection by looking in the direction of the C^α - C^β bond, while 2) is the projection by looking in the direction from the tertiary carbon of the β^3 -hVal side chain towards the C^β .

Kritezer and co-workers studied the influence of side chains in the stabilization of the 14-helix in water by Monte Carlo simulations using a OPLS-AA force field. They found that β^3 side chains that are branched in the first carbon stabilize the 14-helix conformation.¹⁵⁰ The previous authors also report that β^3 -homovaline and β^3 -homoleucine promote the 14-helix, where the first have a more stabilizing effect than the latter.

In the case of the (M)-12-helix β^3 -hVal conformer, it is interesting to notice that the dihedral angles possess opposite signs to the (M)-14-helix although they present the same handedness. This is then reflected in the different orientation of the amide oxygens and hydrogens to the different N and C-terminus, as it was mentioned before. In the (M)-12-helix β^3 -hVal conformer the angles θ present an average value of -89.4 degrees in agreement with the literature.^{99,127} The angles ϕ and ψ present a clinal conformation (between 30 and 150 degrees or between -30 and -150 degrees) which is in good agreement with ones reported in previous studies.^{99,127,149} The solvents effects on the dihedral angles of the (M)-12-helix seem to be rather small. The ϕ and θ decrease by 4 and by 1.4 degrees respectively, while ψ registers a slight increase by 4 degrees in solution. The same trend is observed for the (P)-12-helix. In the case of the 12-helix in both (M) and (P) conformations, the β^3 -hVal side chain presents the same configuration of the in gas phase and in solution. The superposition of the geometries optimized in gas phase and in solution are present in Figure 5.10-a) and Figure 5.10-b for the (M)-12-helix and (P)-12-helix respectively. The small differences in the dihedral angles between gas phase and solution are reflected in a good overlap of the conformations. For the case of the (P)-12-helix the RMSD between the two environments is 0.23 Å, in the case of the (M)-12-helix the RMSD is 3.63 Å. The latter is mainly due to small rotations of the

β^3 -hVal side chains and the cap of the C-terminus.

In order to study the stability of the β^3 -hVal hexapeptides the conformational energies were refined with DFT and wavefunction methods with local approximations. The relative energies related to the most stable conformation are presented in Table 5.9. In gas phase B3LYP, LMP2 and SCS-LMP2 seem to predict that the 14-helix left handed is the most stable conformation. Nonetheless, the difference between the latter conformation and the (P)-12-helix and (M)-12-helix at the LMP2 level is very small, only 0.1 and 0.3 kcal·mol⁻¹ respectively. This leads to believe that the 12-helix could still occur at this level of theory.

Table 5.9: Relative energy differences (in kcal·mol⁻¹) to the most stable conformations of β^3 -hVal hexamers computed with B3LYP and B3LYP-D3 with the cc-pVTZ basis set and with LMP2 and SCS-LMP2 with the aug'-cc-pVTZ basis set. A comparison is made between the 12-helix and the 14-helix both in a left (M) and right (P) helical folding. The structures were optimized with B3LYP-D3/6-31G*. The COSMO model was used to approximate solvent effects for both single point and optimization runs.

	GP				TFE			
	DFT	DFT-D3	LMP2	SCS-LMP2	DFT	DFT-D3	LMP2	SCS-LMP2
(P)-14-H	2.8	5.2	2.7	3.0	2.1	3.1	1.9	2.2
(M)-14-H	0.0	2.9	0.0	0.0	0.0	0.0	0.0	0.0
(P)-12-H	1.0	0.0	0.1	2.1	6.6	1.6	3.7	4.7
(M)-12-H	1.8	0.4	0.3	2.3	6.9	2.1	4.2	5.3

The 12-helices present two hydrogen bonds between two nitrogens of the azide terminal group and the (*i*-2)th residue that maybe contribute to the stability of this conformation. B3LYP with dispersion corrections gives that the preferential conformation is the (P)-12-helix with the right handed structure being destabilized by less than 0.5 kcal·mol⁻¹. When one includes solvent effects, all the methods predict that the preferential adopted conformation is the (M)-14-helix. B3LYP-D3 predicts that the latter in a right helical folding is even more destabilized than the 12-helix in both handedness. All the other methods predict that both 12-helix conformers are less stable than the

(P)-14-helix.

Wu *et al.* reported that the 12-helix is preferred over the 14-helix in gas phase while in methanol (using the SCIPCM model) the stability of the latter dominates.⁹⁹ The stabilization of the 14-helix over the 12-helix in organic solvents like methanol or trifluoroethanol may result from its macrodipole that stabilizes the gauche conformation of θ .¹²⁷ The macrodipole results from the electrostatic interaction between the carbonyl carbon and nitrogen from the amide in the backbone.¹²⁷ Martinek and co-workers also point out that in terms of side chain interactions, the conformation more sensitive to hydrophobic van der Waals interactions is the 14-helix.¹²⁵ The solvent is said to induce attractive interactions between the hydrophobic side chains leading to an increase of the stability of the 14-helix conformation.¹²⁵ Another possible explanation for the higher stability of the 14-helix in solution can be related to the fact that the backbone is more shielded from the solvent. In the 14-helix the hydrogen bond pattern happens between 3 residues, resulting in three residues per turn.⁹⁹ The 12-helix presents around 2.5 residues per turn.⁹⁹ So comparing for example the β^3 -amino acid hexamers in the 14- and 12-helix folding the latter presents a more stretched geometry. The backbone is therefore more accessible to the solvent leading to a possible destabilization of the helix.

The β^3 -hLeu, when compared with β^3 -hVal, presents an extra CH_2 adjacent to the backbone. This allows the side chains to have more rotational degrees of freedom. The optimized structures in trifluoroethanol are shown in Figure 5.12.

Comparing β^3 -hLeu with β^3 -hVal 14-helix one can notice that the side chains are not in atop position as in the case of the latter, but instead they present a rotation of around 15 degrees. This results as a response to the hindrance of the side chains. Contrary to what was observed in the case of the β^3 -hVal, the β^3 -hLeu side chains conformation was retained for the most stable conformations of the 14-helix and 12-helix for gas phase and solution.

Looking into the hydrogen bond lengths presented in Table 5.10, one can notice that the 14-helices present slightly larger bond lengths than the 12-helices. When one goes from the gas phase to solution there is in general a slight decrease of the bond lengths and a slight increase of the angles due to solvation. The hydrogen bonds with the present lengths are known to be characterized as being strong hydrogen bonds.¹⁴⁸

Internal torsional angles are presented in Table 5.11. One can notice that they are very similar to the ones observed for the β^3 -hVal oligomers. For the gas phase geometries, the 14-helices the θ presents a characteristic gauche conformation, while ϕ and ψ are in

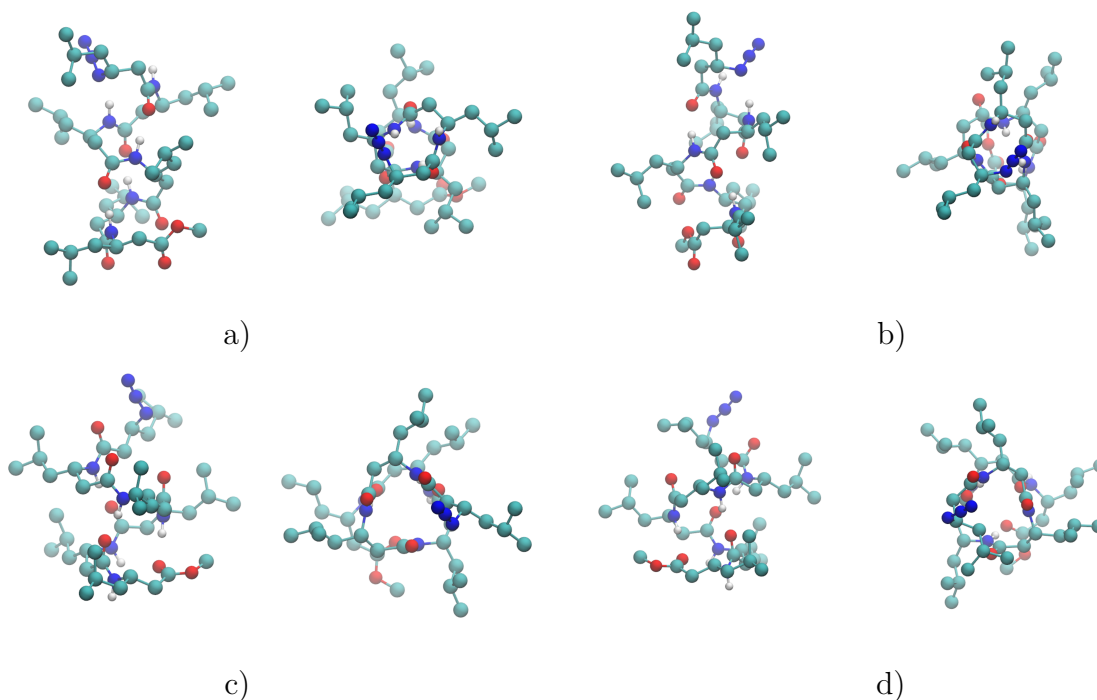


Figure 5.12: Structures of the most stable conformations for the β^3 -hLeu in a 12-helix conformation a) left handed (M), b) right handed (P) and in a 14-helix conformation c) left handed (M) d) right handed (P). The geometries were optimized with B3LYP-D3/6-31G* in trifluoroethanol (TFE). The solvent effects were introduced by using the COSMO model. Aliphatic hydrogens were removed for clarity.

a anticlinal conformation. This angles as it was discussed before for the β^3 -hVal are in agreement with what should be expected for these conformations.

For the case of the 12-helices the θ is very close to the expected 90 or -90 degrees, depending if the helical conformation is right or left handed. The angles ϕ and ψ present a clinal conformation which is in good agreement with the ones reported in the literature.^{99,127,149}

Upon including solvent effects there is a slight decrease of the θ and ψ angles. The ψ angle in the case of the β -peptides with cyclic side chains was shown to be a good indicator of the tightness of the conformation. Taking this into account, one could say that the solvent effects originate a more tight structure, which is reflected by a decrease of the ψ angle.

Superposing the most stable conformations obtained in gas phase and in solution (Figure 5.13), one can see that for 14-helical conformers no significant difference in the geometries is observed. This is also reflected in the small RMSD values of 0.15 Å and of

0.16 Å for the (P)-14-helix and (M)-14-helix respectively. In the case of the 12-helices, a RMSD of 0.91 Å is obtained for the right handed and 1.97 Å for the left handed structures. These larger differences can be a consequence of the decrease of the θ angle from the gas phase to solution. The latter decreases by 8 degrees for the (P)-12-helix and by about 10 degrees for the (M)-12-helix.

Table 5.10: Average hydrogen-bonds lengths (in Å) and angles (in degrees) for β^3 -hLeu in a 12-helix and 14-helix conformations both with a left (M) and right (P) helical folding. The structures were optimized in gas phase (GP) and in trifluoroethanol(TFE) with B3LYP-D3/6-31G*.

	GP		TFE	
	Distance	Angle	Distance	Angle
(P)-14-H	1.998	163.2	1.883	169.0
(M)-14-H	1.987	168.4	1.910	170.3
(P)-12-H	1.953	161.6	1.933	154.2
(M)-12-H	1.950	155.5	1.852	165.5

Table 5.11: Average backbone dihedral angles (in degrees) for β^3 -hLeu in a 12-helix and 14-helix conformations both with a left (M) and right (P) helical folding. The structures were optimized in gas phase (GP) and in trifluoroethanol(TFE) with B3LYP-D3/6-31G*. The torsional degrees of freedom are defined as follows: $\phi = \text{C-N-C}^\beta\text{-C}^\alpha$, $\theta = \text{N-C}^\beta\text{-C}^\alpha\text{-C}$ and $\psi = \text{C}^\beta\text{-C}^\alpha\text{-C-N}$.

	GP			TFE		
	ϕ	θ	ψ	ϕ	θ	ψ
(P)-14-H	137.8	-59.2	140.7	137.8	-58.6	138.2
(M)-14-H	-137.6	59.8	-137.8	-133.3	58.6	-137.4
(P)-12-H	-114.4	92.4	-97.9	-119.2	83.9	-82.7
(M)-12-H	103.2	-93.3	91.1	107.7	-82.4	80.0

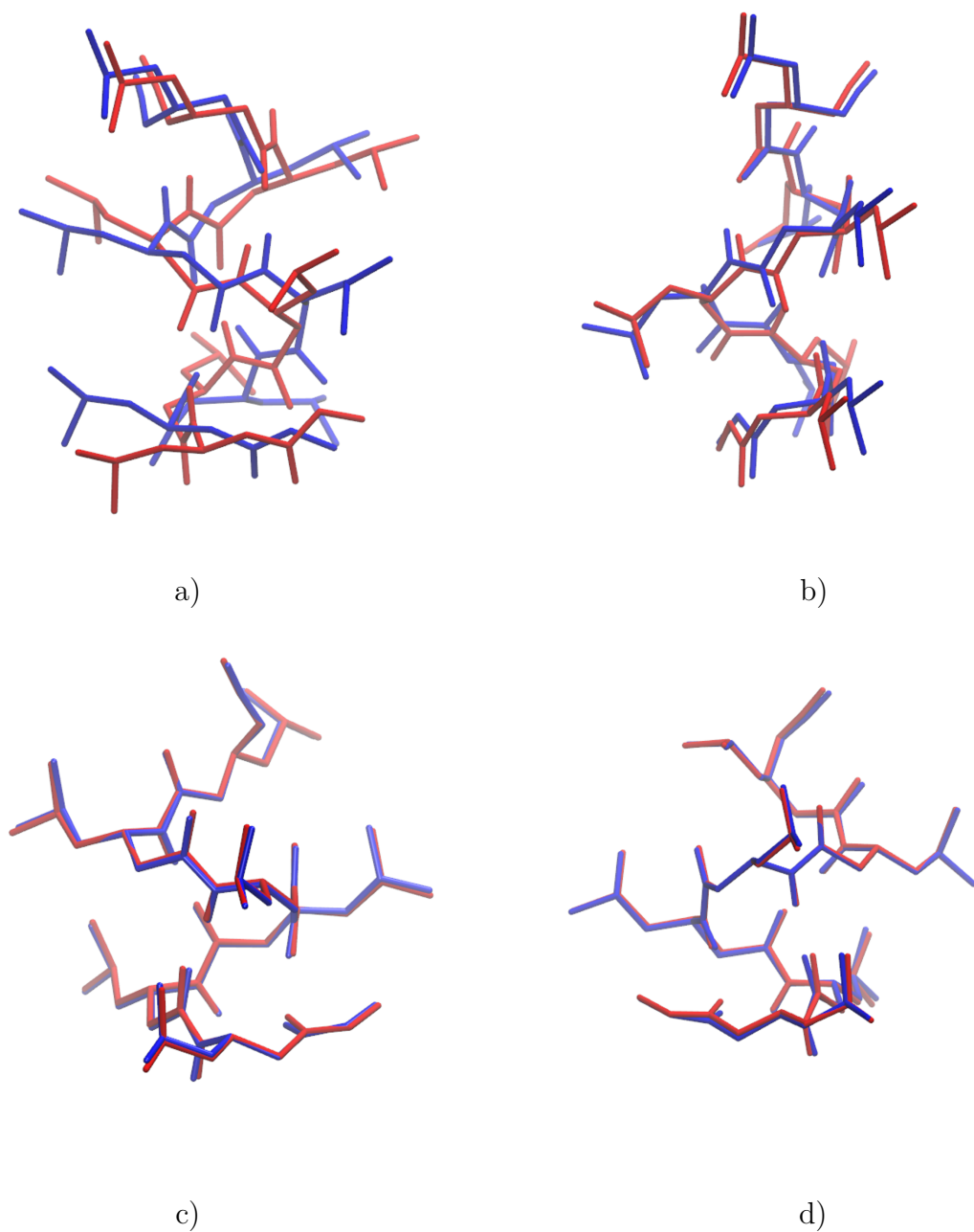


Figure 5.13: Superimposed optimized structures for the β^3 -Leu in a 12-helix conformation a) left handed (M), b) right handed (P) and in a 14-helix conformation c) left handed (M) d) right handed (P). The structures are color coded distinguishing gas phase (blue) and trifluoroethanol (red). Aliphatic hydrogens were removed for clarity.

The conformational stability was assessed in the same way as in the case of the β^3 -hVal. The relative energies to the most stable conformation are present in Table 5.12. Contrary to what was observed for β^3 -hVal hexamers, in gas phase the most stable conformation for the β^3 -hLeu oligomers are the 12-helices. Between the handedness of the latter, B3LYP seems to favour the (M)-12-helix, while the other methods indicate that the most stable conformation is the (P)-12-helix. This disagreement may be related to dispersion forces than are neglected in this method. In the case of B3LYP-D3 and LMP2 the difference to the second most stable conformation is about $2.4 \text{ kcal}\cdot\text{mol}^{-1}$. The stability of the 12-helix in gas phase is in good agreement with what has been reported in the literature, in gas phase the 12-helix dominates over the 14-helix.⁹⁹ Upon inclusion of solvent effects most methods predict that the most stable conformer is the (M)-14-helix. This is in good agreement with the literature, which attributes this stability to the macrodipole of this helical conformation.⁹⁹ B3LYP-D3 predicts the (M)-12-helix as the most stable conformation. However the energy difference for the (M)-14-H and (M)-12-H is quite small, indicating the possible existence of such conformations in solution.

Table 5.12: Relative energy differences (in $\text{kcal}\cdot\text{mol}^{-1}$) to the most stable conformation for β^3 -hLeu hexamers computed at the B3LYP and B3LYP-D3 methods with the cc-pVTZ basis set and with LMP2 and SCS-LMP2 with the aug'-cc-pVTZ basis set. A comparison is made between the 12-helix and the 14-helix both in a left (M) and right (P) helical folding. The structures were optimized with B3LYP-D3/6-31G*. The COSMO model was used to approximate solvent effects for both single point and optimization runs.

	GP				TFE			
	DFT	DFT-D3	LMP2	SCS-LMP2	DFT	DFT-D3	LMP2	SCS-LMP2
(P)-14-H	3.5	2.6	2.4	1.1	6.0	2.1	1.4	2.8
(M)-14-H	0.1	4.1	3.4	0.9	0.0	1.0	0.0	0.0
(P)-12-H	5.0	0.0	0.0	0.0	9.3	0.0	3.9	6.0
(M)-12-H	0.0	2.3	6.4	4.5	4.2	1.0	8.3	9.0

From these results one can conclude that the change of the side chain from β^3 -hVal to β^3 -hLeu does not seem to have influence in the adopted preferential conformation, which is in agreement with Glättli *et al.*¹⁵¹ For the hexamers constituted by only β^3 -hVal or

β^3 -hLeu the most stable conformation in solution seems to be the (M)-14-helix. Raguse *et al.* reported some experimental circular dichroism spectra with very similar systems used in this work.¹⁵² The previous authors observed that β^3 -hVal presented a 14-helix in methanol, while β^3 -hLeu seems to adopt a different conformation, a 12-helix. Nonetheless the authors mention that in order to prove the existence of β^3 -hLeu in a 12-helix, high resolution NMR techniques should be used.¹⁵²

The inclusion of solvent effects leads to a stabilization of the (M)-14-helix over the (P)-12-helix. The 14-helix when compared with the 12-helix presents a more compressed structure along the axis defined by the terminal groups. This is a direct consequence of the hydrogen bond pattern. This results in a larger residue/turn relation, 3.0 for the 14-helix and 2.5 for the 12-helix.⁹⁹ As a consequence, the backbone of the 12-helix is easily accessed by the solvent, which contributes to the destabilization of the peptide folding. This associated with a small macrodipole proposed by Wu *et al.* seem plausible reasons to justify the stabilization of the 14-helix over the 12-helix in solution. This behaviour underlines the importance of including solvent effects when handling this type of systems.

6 Perturbative-Monte Carlo QM/MM

Hybrid QM/MM methods coupled with molecular dynamics (MD) are a widespread methodology for the study of chemical processes in condensed media. In the latter, the total system is separated into two regions, each treated at a different level of theory. One small region is treated with QM while the remaining system (e.g. solvent) is treated with classical force fields. Nevertheless, a large number of electronic structure calculations (wavefunction methods or DFT) are necessary at each MD step. In order to obtain reliable information on the dynamics and thermodynamics of the system, these have to be in the order of hundreds of thousands. This feature makes QM/MM MD computationally demanding, even though most of the degrees of freedom sampled are found in the MM part.¹⁵³ An alternative to the former would be to analyse the MM region by using Monte Carlo. Although the latter is embarrassingly parallel, the need of a QM calculation for every solvent move makes it computationally expensive. In the next section one will present a cost-effective methodology for sampling a system in solution.

6.1 Method

Before going into details, it is important to mention that the method will be discussed in terms of a solute in a aqueous environment. The former will be treated at the QM level while the latter at the MM level. But one should keep in mind that the method by itself can be farther applied for example to system in a proteic surrounding.

Under a QM/MM approach the total Hamiltonian of the system is described by equation (2.66) and one can write the total energy of the system as

$$E_{tot} = \langle \Psi | \hat{H}^{QM} + \sum_i \sum_{\alpha} -\frac{q_{\alpha}}{r_{i\alpha}} | \Psi \rangle + \sum_A \sum_{\alpha} \frac{q_{\alpha} Z_A}{r_{A\alpha}} + E_{s-S}^{vdW} + E_S^{MM}, \quad (6.1)$$

where the indices i and A refer to the solute electrons and nuclei, and α runs over the

solvent atoms. \hat{H}^{QM} stands for the Hamiltonian in vacuum of the QM region and Ψ is the converged wave function. E_{s-S}^{vdW} contains the classical van der Waals interactions described by a 12-6 Lennard-Jones potential between the solute (s) and the solvent (S) and E^{MM} is the classical energy of the solvent. In a Metropolis Monte Carlo simulation when a solvent molecule m moves, the effective Hamiltonian in the first term suffers a change of

$$\Delta\hat{H} = - \sum_i \sum_{\alpha \in m} q_{\alpha} \left(\frac{1}{r'_{i\alpha}} - \frac{1}{r_{i\alpha}} \right), \quad (6.2)$$

where r' is the new position of the solvent molecule.

Truong *et al.*^{15,16} proposed that one could make use of Monte Carlo in a way where the QM and MM sampling are separated. The effect of solvent movements on the QM system could then be treated perturbatively, leading to a low cost simulation method.^{15,16} This methodology will be referred to as perturbative-Monte Carlo QM/MM (PertMC).

In the PertMC one assumes that a random move of a solvent molecule does not change significantly the wavefunction of the solute, in this way one can calculate the effect of this change in the QM energy by using the wavefunction from the previously accepted step.^{15,16} This approach avoids the need to carry out a full SCF calculation in every solvent random move.

In a PertMC, the effect of the lattice change on the wavefunction of the solute is deemed to be small, so that the difference in the energy can be calculated by means of perturbation theory. This avoids a new SCF cycle and only requires trivial nuclear integrals. The change in the total energy is then computed as

$$\Delta E_{tot} = \Delta E_{pert} + \sum_A \sum_{\alpha} q_{\alpha} Z_A \left(\frac{1}{r'_{A\alpha}} - \frac{1}{r_{A\alpha}} \right) + \Delta E_{s-S}^{vdW} + \Delta E^{MM} \quad (6.3)$$

$$\Delta E_{pert} = \sum_{\mu\nu} P_{\mu\nu} \langle \mu | \sum_{\alpha \in m} -q_{\alpha} \left(\frac{1}{r'_{i\alpha}} - \frac{1}{r_{i\alpha}} \right) | \nu \rangle.$$

where ΔE_{pert} is the contribution to the energy due to a perturbative movement of a solvent molecule. $P_{\mu\nu}$ is the density matrix in the AO basis from the previous SCF calculation. The former is calculated in the beginning of the PertMC run and every time one performs an update of the wave function. A schematic representation of the PertMC is presented in Figure 6.1

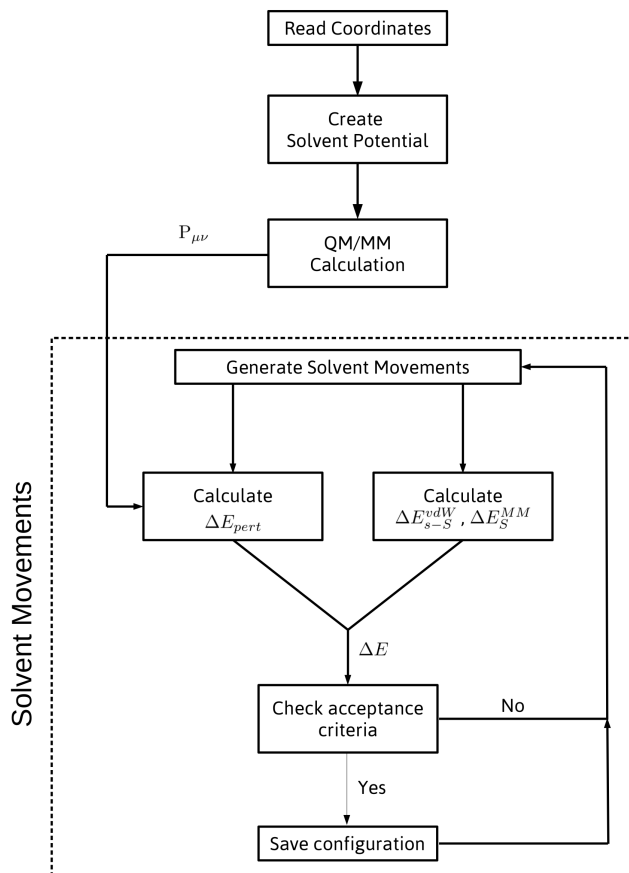


Figure 6.1: Schematic representation of the Perturbative-Monte Carlo QM/MM algorithm.

From equation (6.3), one can see that the main computational cost associated with this procedure is the calculation of the one-electron integrals. In this way it is expected a significant saving in the computational time when compared to a non-perturbative MC approach. The PertMC and MC algorithms were coded in C++ with an interface to the Molpro³⁸ and Orca¹⁴⁴ computational chemistry software packages. The former generates the trial moves and computes the classical energy terms, while the latter programs are responsible for non-classical energy calculations and the ΔE_{pert} term.

In the next subsection one will provide details about the water model used, how the random moves are defined as well as the applied boundary conditions.

6.1.1 Water model

Over the years several intermolecular potentials have been proposed to describe pure water in solution in molecular mechanics. One of these classes is the one proposed by Jorgensen *et al.* designated by TIP n P (Transferable Intermolecular Potential) with n interacting sites.^{154,155} The original models describe the water as non-polarized with a fixed atom structure. Nonetheless there are also flexible and polarized versions of the TIP n P model.^{156,157}

In this work one used the water model TIP3P as in their original formulation.¹⁵⁴ The intermolecular potential (U) between two water molecules, one A and another B is than described as

$$U_{AB} = \sum_i^A \sum_j^B \frac{e^2 q_i q_j}{r_{ij}} + \left(\frac{C}{r_{OO}^{12}} - \frac{D}{r_{OO}^6} \right), \quad (6.4)$$

where q is the partial charge of the atom and r_{ij} is the distance between the two atoms. r_{OO} is the distance between oxygens of the two water molecules and C and D are 12-6 Lennard-Jones parameters relative to the oxygens. The geometry specifications and the parameters for the potential function for a water molecule TIP3P are presented in Table 6.1.

Table 6.1: Geometric and Lennard-Jones parameters for water in the TIP3P model.¹⁵⁴

r(O-H)	α (H-O-H)	$C \times 10^{-3}$	D	q_O	q_H
Å	degrees	kcal Å ¹² ·mol ⁻¹	kcal Å ⁶ ·mol ⁻¹		
0.9572	104.52	629.4	625.5	-0.82	0.41

6.1.2 Translation and rotation movements

Under our implementation, the translation and rotation movements of the solvent molecules are made in relation to their center of mass. The translational movements were implemented as described in section 2.2.1.

In terms of orientational moves for rigid non-linear molecules, like rotation, used was made of quaternions.^{61,62,158,159} The latter are directly connect with the Euler angles.⁶¹

The latter are not so effective to work with, since the rotation operation under this description is associated to trigonometric relations. These are known to be computationally more expensive.

The rotation of a rigid molecule can be related to a quaternion vector (Q). The latter is defined in a four dimensional (4D) space as $Q = (q_0, q_1, q_2, q_3)$ with a unitary norm, $q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1$. One starts by creating a random vector (e_x, e_y, e_z) with unitary norm, $e_x^2 + e_y^2 + e_z^2 = 1$. Then one generates two random numbers ξ_3 and ξ_4 independent and uniform in an interval $[-1, 1]$ until $S = \xi_3^2 + \xi_4^2 < 1$. These new random numbers are different from the ones created for translation. The random vector is defined as in equation (2.64) but in respect to the new random numbers. The components of the Q vector are then defined in relation to this vector as¹⁵⁹

$$\begin{aligned} q_0 &= \cos\left(\frac{\theta}{2}\right) \\ q_1 &= e_x \sin\left(\frac{\theta}{2}\right) \\ q_2 &= e_y \sin\left(\frac{\theta}{2}\right) \\ q_3 &= e_z \sin\left(\frac{\theta}{2}\right), \end{aligned} \tag{6.5}$$

where θ is the rotation angle in radians. The latter is a random number in a uniform distribution between zero and a maximum angle that the molecule is allowed to rotate.

In order to obtain a new set of coordinates, one can represent the latter as a column matrix in 3D space, \mathbf{X}^{new} . The new set of coordinates is then obtained by solving $\mathbf{X}^{\text{new}} = \mathbf{R}\mathbf{X}^{\text{old}}$, where \mathbf{X}^{old} is the matrix that contains the old coordinates and \mathbf{R} is the so called rotation matrix that is defined as^{61,62,158,159}

$$\mathbf{R} = \begin{pmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1q_2 - q_0q_3) & 2(q_1q_3 + q_0q_2) \\ 2(q_1q_2 + q_0q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2q_3 - q_0q_1) \\ 2(q_1q_3 - q_0q_2) & 2(q_2q_3 + q_0q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{pmatrix} \tag{6.6}$$

A complete derivation of such matrix is given in Reference [159].

6.1.3 Simulation of non-periodic systems

Normally MD and MC simulations are carried out using a periodic boundary condition under a minimal image convention,⁶¹ this allows the simulation of a periodic system.

An alternative to the previous approach is the use of a “spheric boundary condition”.^{160,161} This was imposed by a harmonic potential in order to guarantee the system integrity. This is, that the system is maintained at an approximately constant density. The following presents the form

$$V(\mathbf{r}) = \begin{cases} k(\mathbf{r}_i - \mathbf{r}_{sphere})^2 & \text{if } \mathbf{r}_i > \mathbf{r}_{sphere} \\ 0 & \text{if else,} \end{cases} \quad (6.7)$$

where \mathbf{r}_i is the distance between the center of mass of the MM molecule and the center of mass of the system. A graphic of the shape of the potential is presented in Figure 6.2.

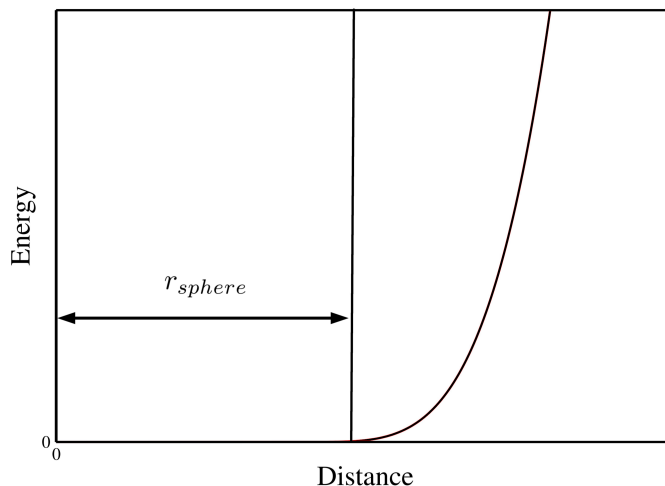


Figure 6.2: Schematic representation of a spheric boundary condition.

One should notice that the potential is only applied when a MM molecule goes beyond the limit of the spheric volume. In this way when one molecule suffers a displacement that puts it in a region of space very far away from the system that configuration will be rejected.

In the next section one present some preliminary results performed with some test systems, where one compares results from a MC QM/MM with a PertMC approach.

6.2 Test systems

Truong *et al.*^{15,16} developed and applied the perturbative-Monte Carlo in a QM/MM approach to study the microsolvation of alkali metals and of halogen anions by water clusters. In the latter study the largest system was composed of 9 water molecules. In a latter studies the first author also report a study of sodium with a 125 water cluster where the QM part was treated with HF and MP2 theory with a 6-31G* basis set.¹⁶² These studies did not include any molecular solute.

As a first approach one is interested in comparing the performance of the perturbative Monte Carlo used in a QM/MM approach with a regular MC run. The test systems are composed of methanol and an arginine amino acid with a small backbone attached in an aqueous surrounding. The water model used was the TIP3P, which presents the characteristics mentioned in the previous section. The QM and respective ΔE_{pert} were calculated using the Orca software package, with the Hartree-Fock with three corrections method (HF-3c).^{144,163}

The first test system is composed of one methanol molecule treated at the QM level, while the MM part is composed by 213 water molecules. In the second system the QM part is composed of an arginine amino acid caped in the same way as in previous sections of this thesis, while the MM part is composed of 581 TIP3P waters. Both QM molecules were optimized at the B3LYP/6-31G* level of theory.

An initial cubic water box was constructed with the TINKER Molecular Mechanics software package,¹⁶⁴ where the QM molecule is placed in the center. The geometry of the latter molecule is fixed and also static during all the simulation, only the solvent is able to perform trial moves.

A spheric boundary condition was imposed with a force constant of $5 \text{ kJ} \cdot \text{mol}^{-1} \text{Å}^{-2}$ for both systems. The value for the force constant was achieved by trial and error. The obtained value was shown to be strong enough to maintain the cohesion of the system and at the same time smooth enough to allow the rearrangement of the water molecules in the outside layer. For the methanol-water cluster the radius of the sphere used was 9 Å, while for the arginine-water cluster was 14 Å. The water molecules in the initial cubic box beyond this distance were deleted.

The system composed of methanol was let to evolved to equilibrium by a simulation run of 8×10^5 steps, followed by a production run with 6×10^5 steps. For the second system the equilibrium was achieved by a simulation run of 1×10^6 steps followed by a

production run with the length of 2×10^6 steps. The simulation was carried out using a NVT ensemble with a temperature of 298 K.

The solvent random steps are composed by simultaneous translation and rotation movements in respect to the molecules center of mass. They were constrained to a maximum displacement and a maximum rotation angle of 0.25 Å and 20 degrees respectively. This was reflected in an acceptance ratio of about 45 % in both systems.

In the PertMC method an update of the solute wavefunction is carried out in an interval of perturbative solvent steps, N_{update} . Simulations were carried out with several N_{update} values namely 250, 500, 1000, 2000 and 4000 for the first system, while for the larger system values of 250, 1000, 4000 and 8000 were used. This allows one to study the influence of the regularity of the wavefunction updates on the energy of the system. It should be noticed that, in the limit of $N_{update} = 1$ the PertMC will correspond to the regular Metropolis Monte Carlo (MC). In every accepted configuration of the PertMC simulation one also performed a full SCF calculation and the results that outcome from the latter are mentioned as regular MC.

One will present a comparison between the average interaction energy of the system using a PertMC and a MC. The former is defined as

$$\begin{aligned}\langle E_{int} \rangle &= \frac{1}{X} \cdot \frac{1}{M_{tot}} \sum_{i=1}^{M_{tot}} [E_{total}(\mathbf{r}_i^N) - E_{QM} - E_{MM}(\mathbf{r}_i^N)] \\ \langle E_{int} \rangle &= \frac{1}{X} \cdot \frac{1}{M_{tot}} \sum_{i=1}^{M_{tot}} E_{int}(\mathbf{r}_i^N),\end{aligned}\tag{6.8}$$

where X is the total number of solvent molecules and M_{tot} is the total number of accepted PertMC configurations. $E_{total}(\mathbf{r}_i^N)$ and $E_{MM}(\mathbf{r}_i^N)$ are the total energy of the system and the energy of the MM part in configuration \mathbf{r}_i^N respectively, while E_{QM} term is the energy of the QM part in gas phase.

One will also assess the difference between the interaction energy obtained by regular MC and by PertMC in terms of absolute values. The former is defined as

$$|E_{err}| = \left| \frac{1}{M_{tot}} \sum_{i=1}^{M_{tot}} [E_{int}^{MC}(\mathbf{r}_i^N) - E_{int}^{PertMC}(\mathbf{r}_i^N)] \right|,\tag{6.9}$$

where $E_{int}^{MC}(\mathbf{r}_i^N)$ and $E_{int}^{PertMC}(\mathbf{r}_i^N)$ are the interaction energies given by MC and by PertMC respectively, in every configuration \mathbf{r}_i^N . It would be also interesting to see how

large the differences between the interaction energy in the MC and the PertMC can be during the simulation run. The maximum value will be represented as $E_{err,MAX}$

One will start by looking to the results for the first test system, composed of methanol solvated by 213 waters. A snapshot from the production run is presented in Figure 6.3.

The results for the interaction energy in relation to the system size as a function of the interval of the wavefunction update for the first test system are presented in Figure 6.4. Each energy value is an average value of four simulation runs, where the standard deviation was found to be quite small, with a maximum value of $29 \text{ J}\cdot\text{mol}^{-1}$ for a $N_{update} = 4000$ in the PertMC approach. The value of $\langle E_{int} \rangle$ presents a variation with the N_{update} . This is as expected, since with the increase of the latter one would predict a deviation from the reference value. The energy difference between the $\langle E_{int} \rangle$ from $N_{update} = 250$ and $N_{update} = 4000$ is lower than $40 \text{ J}\cdot\text{mol}^{-1}$ for the simulated system. It is very interesting to see that comparing the two extremes of N_{update} used, where the latter is 16 times bigger than the first, there is not a drastic difference in the interaction energy. One can also notice that the results from the PertMC and MC correlate quite well.

The difference between the MC and the PertMC interaction energies in term of absolute values, $|E_{err}|$, as a function of N_{update} is presented in Figure 6.5.

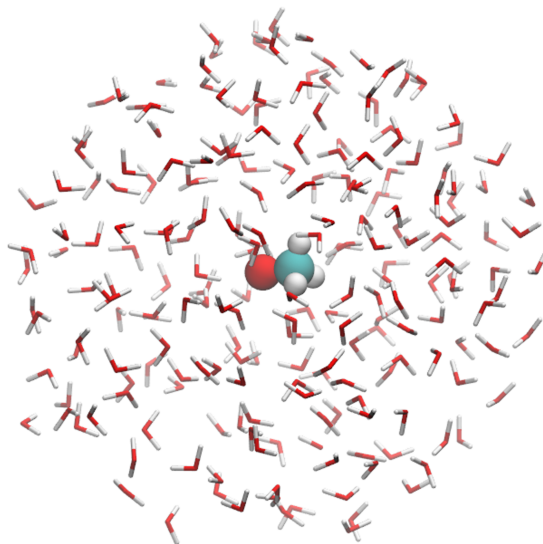


Figure 6.3: System snapshot during the production run of methanol with 213 water molecules TIP3P.

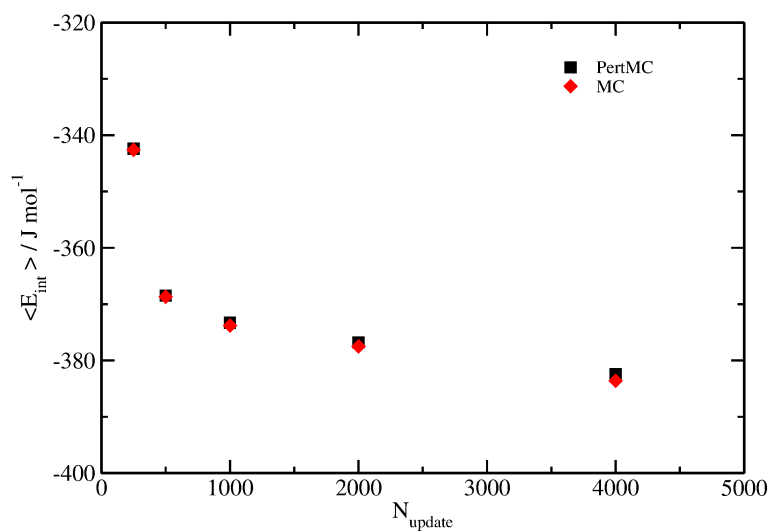


Figure 6.4: Average interaction energy in relation to the system size for MC and PertMC as a function of N_{update} for the methanol-water system.

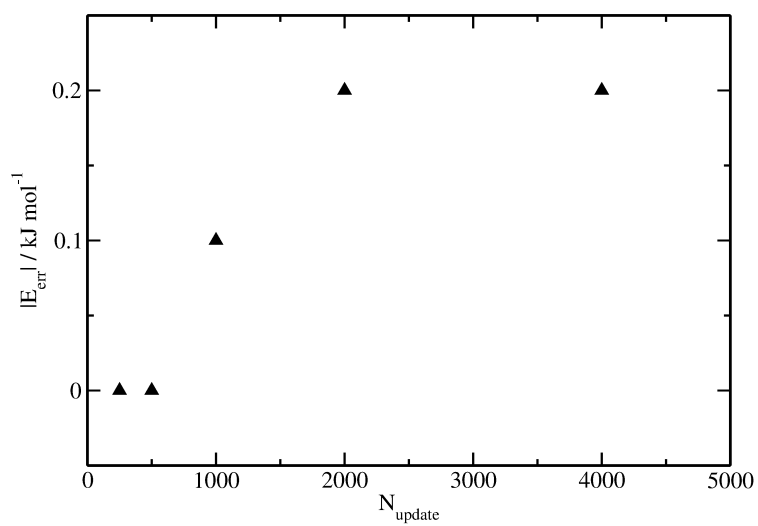


Figure 6.5: Average absolute differences between the interaction energies of the MC and PertMC approaches as a function of N_{update} during the simulation of methanol in aqueous solution.

The $|E_{err}|$ tends to increase with the increase of N_{update} with a maximum difference of $0.2 \text{ kJ}\cdot\text{mol}^{-1}$ for $N_{update} = 4000$. This result is quite relevant, taking into account that one is comparing results where a full SCF cycle is carried out in each accepted structure with values obtained by a perturbative approach. This is an evidence that the assumptions of the PertMC are valid, this is, that not all solvent movements affect significantly the wavefunction of the solute.

The maximum energy differences, $E_{err,MAX}$, between the MC and PertMC during the simulation run as a function of N_{update} are presented in Figure 6.6. $E_{err,MAX}$ tend to increase with N_{update} . The former presents values in an interval of 0.5 to $2.0 \text{ kJ}\cdot\text{mol}^{-1}$, which correspond to a N_{update} of 250 and 4000 respectively. The latter presents the highest standard deviation of $0.8 \text{ kJ}\cdot\text{mol}^{-1}$. This seems to indicate that the perturbative approach is quite a robust method, since there is no huge discrepancy when compared with the non-perturbative approach.

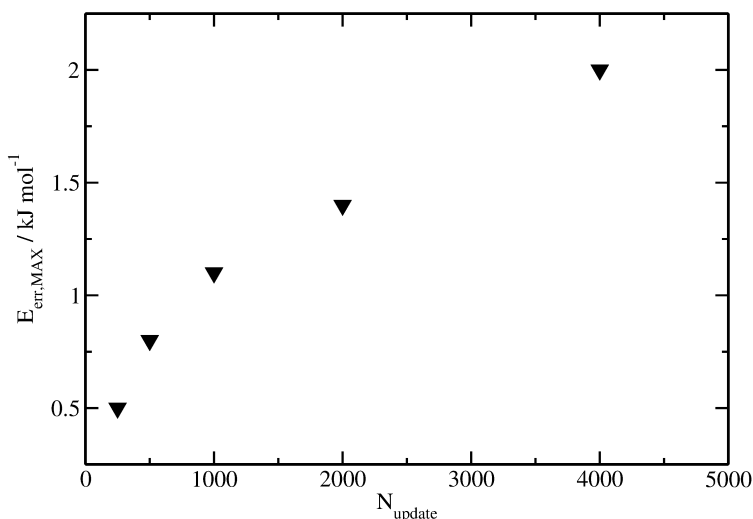


Figure 6.6: Maximum absolute difference between MC and PertMC approaches during the simulation run as a function of N_{update} during the simulation of methanol in aqueous solution.

Now one will take a closer look to the second test system, composed of an arginine model with 281 waters TIP3P. A snapshot from the production run is presented in Figure 6.7.

The average interaction energy in relation to the system size as a function of the interval of the wavefunction update are presented in Figure 6.8.

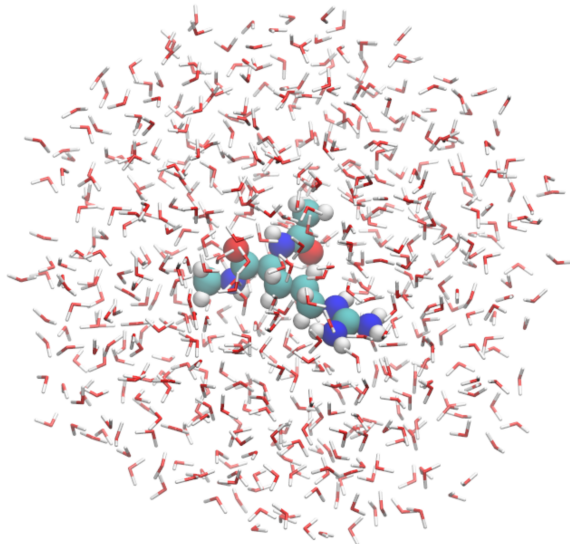


Figure 6.7: System snapshot during the production run of arginine with 581 water molecules TIP3P.

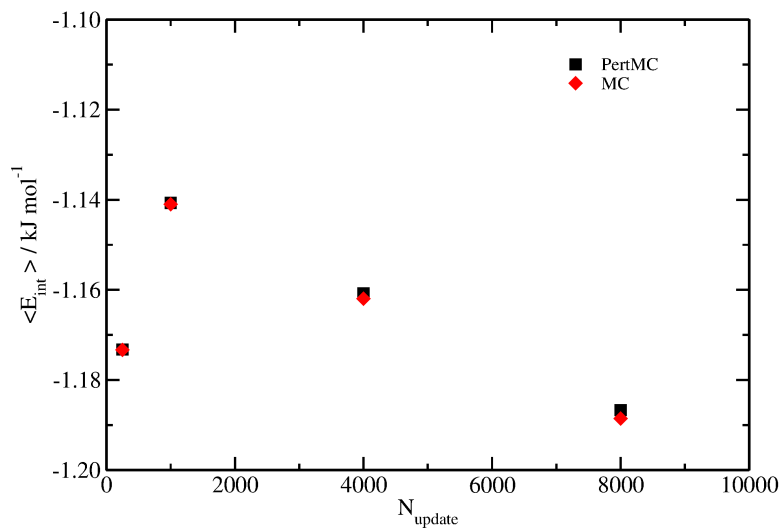


Figure 6.8: Average interaction energy in relation to the system size for MC and PertMC as a function of N_{update} for the arginine-water system.

As it was observed for the first system $|E_{err}|$ tends to increase with the increase of N_{update} , Figure 6.9, with a maximum difference of $1.1 \text{ kJ}\cdot\text{mol}^{-1}$ for $N_{update} = 8000$. Taking into account the system size, this seems to imply once again that not all solvent movements affect significantly the wavefunction of the solute.

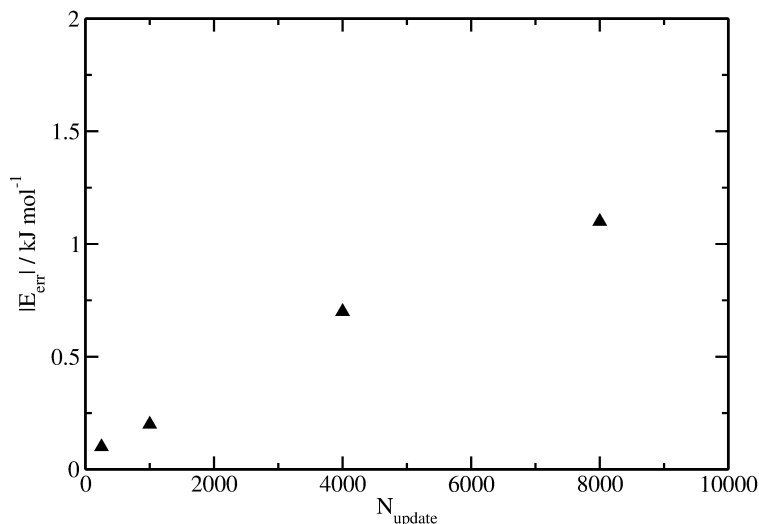


Figure 6.9: Average absolute differences between the interaction energies of the MC and PertMC approaches as a function of N_{update} during the simulation of arginine in aqueous solution.

The $E_{err,MAX}$ between the MC and PertMC during the simulation run as a function of N_{update} are presented in Figure 6.10. As it was observed for the methanol system the $E_{err,MAX}$ tends to increase with N_{update} . The former presents values in an interval of 0.1 to $4.3 \text{ kJ}\cdot\text{mol}^{-1}$, which correspond to a N_{update} of 250 and 8000 respectively. The latter presents the highest standard deviation of $0.7 \text{ kJ}\cdot\text{mol}^{-1}$.

Comparing the $E_{err,MAX}$ at $N_{update} = 4000$ for the two test systems it is surprising to see that the $E_{err,MAX}$ registered are quite close. the difference between the latter are about $1 \text{ kJ}\cdot\text{mol}^{-1}$. This seems to indicate that $E_{err,MAX}$ is independent of system, but instead dependent of the N_{update} .

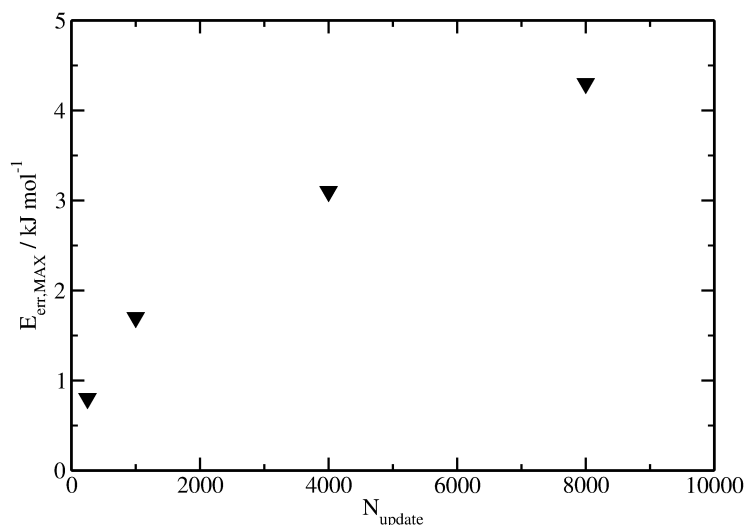


Figure 6.10: Maximum absolute difference between MC and PertMC approaches during the simulation run as a function of N_{update} during the simulation of arginine in aqueous solution.

In general, one would expect that solvent movements that have a higher impact in the wavefunction of the QM molecule are those that take place in the vicinity of the latter. This argument could also be explored in the future in order to regulate the periodicity of the N_{update} .

Although the results presented in this section are preliminary, one showed that PertMC can be applied to more complex systems than the ones studied by Truong *et al.*. These results are quite promising, since the differences between the MC and PertMC approaches are rather small. In this way using the perturbative approach in QM/MM calculations will result in significant computational savings. One should also have in mind that this is still an ongoing project. In this way the assumptions here taken, as for example the fixed and static geometry of the methanol, are valid only for terms of comparison between the MC and PertMC, and to prove the legitimacy to the latter.

7 Conclusions

The main objective of this work was to explore the effects that lead peptides to adopt a certain preferential conformations in solution. Taking this into account one directed our attention to the study of the stability of peptides conformations in solution, as well as to the study of the forces behind their stability in ionic solutions.

On the first work of this manuscript one presented LMP2 calculations in combination with the COSMO model. Use was made of analytical energy gradients under a non-interactive energy only scheme. The method was applied to the optimization of several systems composed of monomers and dimers of small molecules to more complex systems in solution. The simulated solvents were toluene and water, since they represent two very different polarity environments. The degrees of freedom of the studied test set were examined and compared with the canonical method COSMO-MP2. One verified that the effects on the molecular geometries due to local approximations were rather small for both continuum media.

It was observed that the solvent presents a strong effect in the geometries of charged species as in the case of amino acid-halides, this was more evident for protonated arginine-chloride. Upon the inclusion of solvent effects, the interaction energy of such systems was reduced by one order of magnitude, when compared with the gas phase. These observations highlight the importance of including solvent effects when studying such interactions, even if through an implicit solvation model.

The COSMO-LMP2 approach demonstrated to be a cost-effective method for the study of molecular systems in solution when compared to the canonical approach. The former exhibits a reduced computational scaling and the advantage of neglecting by construction the basis set superposition effects in the correlation energy. These features make it a suitable method to be applied to systems with over 50 atoms. This was demonstrated by applying such procedure to study the conformational stability of a small β -peptide constituted by 74 atoms with quite promising results.

In the second work, PES for a series of different α -amino acids model systems in the

presence of halide anions were calculated with COSMO-LMP2. The calculated PES demonstrated that there is an increase in the affinity from iodide to fluoride for the charged residues. In the case of bromide and iodide, the PES exhibit a quite flat surface around the amino acid moiety. This suggests that these anions present a larger mobility than chloride and fluoride. This leads one to believe that they present the ability to interact with hydrophobic regions within the peptide. This is confirmed by the analysis of the anions interaction with a backbone model and with the π -systems from the amino acid moieties.

Our results are shown to be in agreement with the observations from Heyda *et al.* in a molecular dynamics study with similar systems.¹⁰⁸ The previous authors mapped the anion density around the amino acid moieties, and verified that fluoride has a high affinity towards the charged group of the amino acid. Such density was shown to be more diffuse along the studied anionic series. In the case of iodide, the density was likely distributed over the amino acids, including hydrophobic areas of the latter. Although our methodology is based in a total different approach, it is in quite surprising agreement with the one from Heyda *et al.*. This observation contributes to the validation of our approach towards the treatment of such specific amino acid-ion effects.

The use of COSMO-LMP2 presents the advantage of allowing the partition of the interaction energy. This allows one to gain access to the contribution of dispersion forces in the latter. The interaction energy profiles for iodide and the protonated histidine demonstrated that such forces contribute with about 50 % for the interaction energy at the sharpest minimum. For the case of fluoride, on the other hand, this contribution is rather small, only about 5 %. This outlines the importance of dispersion forces in such systems.

If one also considers the observations from Chapter 3, one would expect a competition of the large anions with water for the amino acid charged groups, which should also promote the interaction of the former with hydrophobic regions of the amino acid. For fluoride, such effect is not expected, due to strong electrostatic interactions with the charged regions of the amino acid.

One can map the anionic series used in this work with the Hofmeister's anionic salting-out series. In the latter, the fluoride is known as salting-out anion. One can attribute such behaviour to the propensity of this anion to specifically interact with the hydrophilic areas of the peptide. This would allow the backbone of the protein to present enough malleability to fold into a certain conformation. On the other end, iodide is known

as a salting-in anion. This effect may be related to the possibility of iodide to visit hydrophobic regions near the protein backbone, promoting a hindrance effect. The latter might possibly provoke constraints in the backbone preventing the folding of the peptide. This would contribute to the destabilization of the latter, increasing its solubility.

In the third work one applied COSMO-LMP2 to conformational studies of β -peptides with cyclic and aliphatic side chains. The β -peptides with cyclic side chains were constituted of hexamers of homo-oligomers of *cis*-ACPC and *cis*-FAA. It was observed that the first presents a sheet-like structure, while the latter adopts a distinct 14-helix with a left-handed helicity. These observations are in agreement with circular dichroism spectra in TFE and NOESY experiments in deuterated chloroform performed by Pandey and coworkers.¹⁴¹

When compared with *cis*-ACPC, *cis*-FAA presents a CH₂ group replaced by an oxygen. The *cis*-FAA hexamer proved to have a more compact conformation than *cis*-ACPC, which favours close dispersion contacts. This was evidenced by comparing COSMO-B3LYP with COSMO-LMP2 results, in which the former lacks the description of van der Waals forces. Dispersion corrections, when applied to DFT energies, greatly compensates the short-coming of the functional. The more compact packing of the *cis*-FAA is found to be due to a more favourable interaction between the ring and the backbone of the peptide. This occurs namely between the amide hydrogen from the peptidic bond that points towards the oxygen of the side chain. On the *cis*-ACPC this electrostatic interaction is absent and steric repulsion dominates.

The study involving β -peptides with aliphatic side chains involved hexamers of β^3 -hVal to β^3 -hLeu in a 12- and 14-helix conformations both with a left and right handedness. From our results, one could conclude that the change of the side chain from β^3 -hVal to β^3 -hLeu does not seem to have an influence in the adopted preferential conformation. This is in agreement with the work of Glättli *et al.*¹⁵¹ It was shown that both β^3 -hVal and β^3 -hLeu adopt preferentially a (M)-14-helix. The stabilization of the latter over the 12-helix may be related to a more compact structure. This results as a direct consequence of the hydrogen-bonding pattern. In the 14-helix each turn presents a larger number of residues per turn. As a consequence of such structural characteristic, the side chains can contribute to shield the backbone from the solvent and therefore to an increase of the stability of the conformation. The latter effect associated with a small macrodipole proposed by Wu *et al.* seem plausible reasons to justify the stabilization of the 14-helix over the 12-helix in solution.

The use of local correlation methods to the study of β -peptides holds good promise, as one should be able to benchmark the conformational properties of small chains, just as previously done in the case of α -peptides. The first are particularly challenging, due to the large number of energetically close lying conformers which have to be analysed.

The last Chapter of this manuscript was dedicated to a Perturbative Monte Carlo QM/MM scheme. Under this approach the steps in the MM system are handled perturbatively. The code is responsible for generating trial movements in the MM part and the computation of the classical energy terms. The algorithm presents an interface to Molpro³⁸ and Orca¹⁴⁴ software packages, which are responsible for the calculation of the non-classical terms to the total energy.

The main advantage of a pertMC approach is to avoid the need to perform a full SCF calculation per MM step. This, by obvious reasons, leads to a decrease in the computational effort. As a first set of preliminary studies, one applied this approach to the study of the interaction energy using two clusters. One composed of a methanol and another composed of an arginine amino acid with a small section of a backbone in aqueous medium. Several QM wavefunction updates intervals were explored. In every accepted configuration from the PertMC one also performed a full SCF single point in order to evaluate the error from the perturbative approach.

In general, the error between the perturbative and non-perturbative approaches revealed to be rather small, increasing with the increase of the interval of the QM wavefunction update. This was shown even for the cases where the wavefunction update was up to some thousand perturbative steps. This seems to reinforce the idea that some MM perturbative movements do not have a significant impact in the wavefunction of the QM part. One should have in mind that one made use of a static QM conformation. Nonetheless, the obtained results, demonstrate that the use of a perturbative Monte Carlo in QM/MM is quite promising.

Normally in a Monte Carlo QM/MM scheme, one needs to perform a QM calculation per MM step. The latter are usually around several millions in order to have a good statistic. The use of a perturbative approach can easily reduce the number of QM calculations by a factor of 1×10^3 . Although this can be reflected in significant computational savings, one is still limited at the QM level. In the latter, the size of the system is still a delimiting factor. One could think that in a future work, it would be interesting for example to explore the potential of local correlation methods, such as LMP2, within a PertMC approach. Another possibility would be to apply perturbative Monte Carlo into

a QM/MM/CSM scheme. This would allow a decrease of the number of explicit solvent molecules and at the same time compensate the treatment of solute-solvent specific interactions, which is deficient in CSMs.

Bibliography

- (1) A. W. Omta, M. F. Kropman, S. Woutersen and H. J. Bakker, *Science*, July 2003, **301**, 347–349.
- (2) J. D. Batchelor, A. Olteanu, A. Tripathy and G. J. Pielak, *Journal of the American Chemical Society*, Feb. 2004, **126**, 1958–1961.
- (3) J. D. Smith, R. J. Saykally and P. L. Geissler, *Journal of the American Chemical Society*, Nov. 2007, **129**, 13847–13856.
- (4) J. Tomasi and M. Persico, *Chemical Reviews*, Nov. 1994, **94**, 2027–2094.
- (5) J. Tomasi, B. Mennucci and R. Cammi, *Chemical Reviews*, Aug. 2005, **105**, 2999–3093.
- (6) A. Klamt and G. Schüürmann, *Journal of the Chemical Society, Perkin Transactions 2*, 1993, 799–805.
- (7) A. Klamt and V. Jonas, *The Journal of Chemical Physics*, 1996, **105**, 9972–9981.
- (8) S. Miertuš, E. Scrocco and J. Tomasi, *Chemical Physics*, Feb. 1981, **55**, 117–129.
- (9) R. Cammi and J. Tomasi, *Journal of Computational Chemistry*, Dec. 1995, **16**, 1449–1458.
- (10) F. Floris and J. Tomasi, *Journal of Computational Chemistry*, July 1989, **10**, 616–627.
- (11) F. Claeysens, J. N. Harvey, F. R. Manby, R. A. Mata, A. J. Mulholland, K. E. Ranaghan, M. Schütz, S. Thiel, W. Thiel and H.-J. Werner, *Angewandte Chemie*, Oct. 2006, **118**, 7010–7013.
- (12) R. A. Mata, H.-J. Werner, S. Thiel and W. Thiel, *The Journal of Chemical Physics*, Jan. 2008, **128**, 025104.
- (13) R. A. Mata, *Physical Chemistry Chemical Physics*, May 2010, **12**, 5041–5052.

-
- (14) J. M. Dieterich, H.-J. Werner, R. A. Mata, S. Metz and W. Thiel, *The Journal of Chemical Physics*, Jan. 2010, **132**, 035101.
- (15) T. Truong and E. Stefanovich, *Chemical Physics Letters*, 1996, **4**, 348–352.
- (16) T. Truong and E. Stefanovich, *Chemical Physics*, May 1997, **218**, 31–36.
- (17) P. Pulay, *Chemical Physics Letters*, Sept. 1983, **100**, 151–154.
- (18) S. Saebø and P. Pulay, *Annual Review of Physical Chemistry*, 1993, **44**, 213–236.
- (19) C. Hampel and H.-J. Werner, *The Journal of Chemical Physics*, 1996, **104**, 6286–6297.
- (20) M. Schütz, G. Hetzer and H.-J. Werner, *The Journal of Chemical Physics*, 1999, **111**, 5691–5705.
- (21) M. Schütz and H.-J. Werner, *Chemical Physics Letters*, Feb. 2000, **318**, 370–378.
- (22) T. Korona and H.-J. Werner, *The Journal of Chemical Physics*, 2003, **118**, 3006–3019.
- (23) D. Kats, T. Korona and M. Schütz, *The Journal of Chemical Physics*, Sept. 2006, **125**, 104106.
- (24) D. Kats, T. Korona and M. Schütz, *The Journal of Chemical Physics*, Aug. 2007, **127**, 064107.
- (25) D. B. Cook, *Handbook of Computational Quantum Chemistry*, Dover Publications, Mineola, NY, 2005, p. 832.
- (26) C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models*, Wiley, West Sussex, 2nd Ed., 2004, p. 618.
- (27) F. Jensen, *Introduction to Computational Chemistry*, Wiley, West Sussex, 2nd Ed., 2006, p. 620.
- (28) P. W. Atkins and R. S. Friedman, *Molecular Quantum Mechanics*, Oxford University Press, Oxford, 3rd Ed., 1999, p. 545.
- (29) A. Szabo and N. S. Ostlun, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Dover Publications, Mineola, NY, 1996, p. 480.
- (30) K. Raghavachari and J. B. Anderson, *The Journal of Physical Chemistry*, Jan. 1996, **100**, 12960–12973.

- (31) P. Knowles, H.-J. Werner and M. Schütz, *Modern Methods and Algorithms of Quantum Chemistry*, ed. NIC-Directors, Jülich, 2nd Ed., 2000, vol. 3, pp. 97–161.
- (32) S. Saebø and P. Pulay, *The Journal of Chemical Physics*, 1988, **88**, 1884–1890.
- (33) P. Pulay and S. Saebø, *Theoretica Chimica Acta*, June 1986, **69**, 357–368.
- (34) J. Pipek and P. G. Mezey, *The Journal of Chemical Physics*, 1989, **90**, 4916–4926.
- (35) J. W. Boughton and P. Pulay, *Journal of Computational Chemistry*, 1993, **14**, 736–740.
- (36) C. Edmiston and K. Ruedenberg, *Reviews of Modern Physics*, July 1963, **35**, 457–464.
- (37) R. A. Mata and H.-J. Werner, *Journal of Molecular Physics*, Oct. 2007, **105**, 2753–2761.
- (38) H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson and M. Wang, *MOLPRO, version 2012.1, a package of ab initio programs*, see <http://www.molpro.net>, Cardiff, UK, 2012.
- (39) G. Rauhut, P. Pulay and H.-J. Werner, *Journal of Computational Chemistry*, Aug. 1998, **19**, 1241–1254.
- (40) G. Hetzer, M. Schütz, H. Stoll and H.-J. Werner, *The Journal of Chemical Physics*, 2000, **113**, 9443–9455.
- (41) A. El Azhary, G. Rauhut, P. Pulay and H.-J. Werner, *The Journal of Chemical Physics*, 1998, **108**, 5185–5193.
- (42) M. Schütz, H.-J. Werner, R. Lindh and F. R. Manby, *The Journal of Chemical Physics*, July 2004, **121**, 737–750.

-
- (43) H.-J. Werner, F. R. Manby and P. J. Knowles, *The Journal of Chemical Physics*, 2003, **118**, 8149–8160.
- (44) M. Schütz, G. Rauhut and H.-J. Werner, *The Journal of Physical Chemistry A*, July 1998, **102**, 5997–6003.
- (45) A. J. Cohen, P. Mori-Sánchez and W. Yang, *Chemical Reviews*, Jan. 2012, **112**, 289–320.
- (46) S. H. Vosko, L. Wilk and M. Nusair, *Canadian Journal of Physics*, Aug. 1980, **58**, 1200–1211.
- (47) J. P. Perdew and Y. Wang, *Physical Review B*, June 1992, **45**, 13244–13249.
- (48) A. D. Becke, *Physical Review A*, Sept. 1988, **38**, 3098–3100.
- (49) J. P. Perdew, K. Burke and M. Ernzerhof, *Physical Review Letters*, Oct. 1996, **77**, 3865–3868.
- (50) C. Lee, W. Yang and R. G. Parr, *Physical Review B*, Jan. 1988, **37**, 785–789.
- (51) A. D. Becke, *The Journal of Chemical Physics*, 1993, **98**, 5648–5652.
- (52) J. Slater, *Physical Review*, Feb. 1951, **81**, 385–390.
- (53) S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *The Journal of Chemical Physics*, Apr. 2010, **132**, 154104.
- (54) S. Grimme, S. Ehrlich and L. Goerigk, *Journal of Computational Chemistry*, May 2011, **32**, 1456–1465.
- (55) S. Grimme, *Chemistry – A European Journal*, Aug. 2012, **18**, 9955–9964.
- (56) W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *Journal of the American Chemical Society*, Jan. 1996, **118**, 11225–11236.
- (57) W. L. Jorgensen and J. Tirado-rives, *Journal of the American Chemical Society*, 1988, **110**, 1657–1666.
- (58) M. C. Owen, L. Tóth, B. Jojárt, I. Komáromi, I. G. Csizmadia and B. Viskolcz, *Journal of Chemical Theory and Computation*, Aug. 2012, **8**, 2569–2580.
- (59) P. Hünenberger, *Advanced Computer Simulation*, 2005, 105–149.
- (60) N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, *The Journal of Chemical Physics*, 1953, **21**, 1087–1092.

- (61) D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, Academic Press, San Diego, California, 2nd Ed., 2001, p. 638.
- (62) F. J. Vesely, *Journal of Computational Physics*, Aug. 1982, **47**, 291–296.
- (63) M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, 1989, p. 385.
- (64) P. A. Bash, M. J. Field and M. Karplus, *Journal of the American Chemical Society*, Dec. 1987, **109**, 8092–8094.
- (65) S. Re and K. Morokuma, *The Journal of Physical Chemistry A*, Aug. 2001, **105**, 7185–7197.
- (66) T. Vreven, K. S. Byun, I. Komáromi, S. Dapprich, J. A. Montgomery, K. Morokuma and M. J. Frisch, *Journal of Chemical Theory and Computation*, May 2006, **2**, 815–826.
- (67) X. Wang, J. Liu, J. Z. H. Zhang and X. He, *The Journal of Physical Chemistry A*, Aug. 2013, **117**, 7149–7161.
- (68) S. Camuto, *Solvation Effects on Molecules and Biomolecules: Computational Methods and Applications*, Springer, 2008, p. 536.
- (69) B. Mennucci, C. Cappelli, R. Cammi and J. Tomasi, *Chirality*, 2011, **23**, 717–729.
- (70) R. Cammi, B. Mennucci and J. Tomasi, *The Journal of Physical Chemistry A*, Nov. 1999, **103**, 9100–9108.
- (71) M. Cossi, G. Scalmani, N. Rega and V. Barone, *The Journal of Chemical Physics*, 2002, **117**, 43–54.
- (72) B. Mennucci, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, May 2012, **2**, 386–404.
- (73) B. Mennucci and R. Cammi, *Continuum Solvation Models in Chemical Physics*, John Wiley & Sons, Ltd, Chichester, UK, Nov. 2007, p. 619.
- (74) D. M. Dolney, G. D. Hawkins, P. Winget, D. A. Liotard, C. J. Cramer and D. G. Truhlar, *Journal of Computational Chemistry*, Apr. 2000, **21**, 340–366.
- (75) V. Weijs, B. Mennucci and L. Frediani, *Journal of Chemical Theory and Computation*, Nov. 2010, **6**, 3358–3364.

-
- (76) B. Mennucci, *The Journal of Physical Chemistry Letters*, May 2010, **1**, 1666–1674.
- (77) A. Klamt, B. Mennucci, J. Tomasi, V. Barone, C. Curutchet, M. Orozco and F. J. Luque, *Accounts of Chemical Research*, Apr. 2009, **42**, 489–492.
- (78) D. W. Tondo and J. R. Pliego, *The Journal of Physical Chemistry A*, Jan. 2005, **109**, 507–511.
- (79) J. R. Pliego and J. M. Riveros, *The Journal of Physical Chemistry A*, Aug. 2001, **105**, 7241–7247.
- (80) J. R. Pliego and J. M. Riveros, *The Journal of Physical Chemistry A*, Aug. 2002, **106**, 7434–7439.
- (81) F. del Valle and J. Tomasi, *Chemical Physics*, Feb. 1991, **150**, 139–150.
- (82) M. Aguilar, F. del Valle and J. Tomasi, *Chemical Physics*, Feb. 1991, **150**, 151–161.
- (83) J. G. Ángyán, *Chemical Physics Letters*, July 1995, **241**, 51–56.
- (84) C. J. Cramer and D. G. Truhlar, *Chemical Reviews*, Aug. 1999, **99**, 2161–2200.
- (85) O. Christiansen and K. V. Mikkelsen, *The Journal of Chemical Physics*, 1999, **110**, 1365–1375.
- (86) F. Lipparini, G. Scalmani and B. Mennucci, *Physical Chemistry Chemical Physics*, Dec. 2009, **11**, 11617–23.
- (87) R. Cammi, *The Journal of Chemical Physics*, Oct. 2009, **131**, 164104.
- (88) M. Caricato, G. Scalmani, G. W. Trucks and M. J. Frisch, *The Journal of Physical Chemistry Letters*, Aug. 2010, **1**, 2369–2373.
- (89) J. G. Ángyán, *International Journal of Quantum Chemistry*, Sept. 1993, **47**, 469–483.
- (90) T. H. Dunning, *The Journal of Chemical Physics*, 1989, **90**, 1007–1023.
- (91) J. M. Dieterich, J. C. A. Oliveira and R. A. Mata, *Journal of Chemical Theory and Computation*, Sept. 2012, **8**, 3053–3060.
- (92) R. Polly, H.-J. Werner, F. R. Manby and P. J. Knowles, *Journal of Molecular Physics*, Nov. 2004, **102**, 2311–2321.

-
- (93) F. Weigend, A. Köhn and C. Hättig, *The Journal of Chemical Physics*, 2002, **116**, 3175–3183.
- (94) R. A. Mata and H.-J. Werner, *The Journal of Chemical Physics*, Nov. 2006, **125**, 184110.
- (95) I. Rozas and P. E. Kruger, *Journal of Chemical Theory and Computation*, Sept. 2005, **1**, 1055–1062.
- (96) I. Rozas, I. Alkorta and J. Elguero, *Structural Chemistry*, Sept. 2008, **19**, 923–933.
- (97) M. D. Best, S. L. Tobey and E. V. Anslyn, *Coordination Chemistry Reviews*, May 2003, **240**, 3–15.
- (98) J. A. Montgomery, M. J. Frisch, J. W. Ochterski and G. A. Petersson, *The Journal of Chemical Physics*, 2000, **112**, 6532–6542.
- (99) Y.-D. Wu, W. Han, D.-P. Wang, Y. Gao and Y.-L. Zhao, *Accounts of Chemical Research*, Oct. 2008, **41**, 1418–1427.
- (100) T. A. Martinek, I. M. Mándity, L. Fülöp, G. K. Tóth, E. Vass, M. Hollósi, E. Forró and F. Fülöp, *Journal of the American Chemical Society*, Oct. 2006, **128**, 13539–13544.
- (101) J. Kaminsky, R. A. Mata, H.-J. Werner and F. Jensen, *Molecular Physics*, Aug. 2008, **106**, 1899–1906.
- (102) Y. Zhang and P. S. Cremer, *Current Opinion in Chemical Biology*, Dec. 2006, **10**, 658–663.
- (103) W. Kunz, J. Henle and B. W. Ninham, *Current Opinion in Colloid & Interface Science*, 2004, **9**, 19–37.
- (104) R. L. Baldwin, *Biophysical Journal*, 1996, **71**, 2056–2063.
- (105) Y. Zhang and P. S. Cremer, *Annual Review of Physical Chemistry*, Jan. 2010, **61**, 63–83.
- (106) P. Jungwirth and B. Winter, *Annual Review of Physical Chemistry*, Jan. 2008, **59**, 343–366.

- (107) J. Paterová, K. B. Rembert, J. Heyda, Y. Kurra, H. I. Okur, W. R. Liu, C. Hilty, P. S. Cremer and P. Jungwirth, *The Journal of Physical Chemistry B*, July 2013, **117**, 8150–8158.
- (108) J. Heyda, T. Hrobárik and P. Jungwirth, *The Journal of Physical Chemistry A*, Mar. 2009, **113**, 1969–1975.
- (109) K. A. Peterson, D. Figgen, E. Goll, H. Stoll and M. Dolg, *The Journal of Chemical Physics*, 2003, **119**, 11113–11123.
- (110) K. A. Peterson, B. C. Shepler, D. Figgen and H. Stoll, *The Journal of Physical Chemistry A*, Dec. 2006, **110**, 13877–13883.
- (111) F. Weigend, *Journal of Computational Chemistry*, Jan. 2008, **29**, 167–175.
- (112) C. Hättig, *Physical Chemistry Chemical Physics*, 2005, **7**, 59–66.
- (113) M. Boström, D. R. M. Williams and B. W. Ninham, *Langmuir*, July 2001, **17**, 4475–4478.
- (114) L. Vrbka, M. Mucha, B. Minofar, P. Jungwirth, E. C. Brown and D. J. Tobias, *Current Opinion in Colloid & Interface Science*, Aug. 2004, **9**, 67–73.
- (115) A. Frontera, D. Quiñonero and P. M. Deyà, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, May 2011, **1**, 440–459.
- (116) A. Frontera, P. Gamez, M. Mascal, T. J. Mooibroek and J. Reedijk, *Angewandte Chemie (International Ed. in English)*, Oct. 2011, **50**, 9564–9583.
- (117) A. Garcia-Raso, F. M. Albertí, J. J. Fiol, A. Tasada, M. Barceló-Oliver, E. Molins, D. Escudero, A. Frontera, D. Quiñonero and P. M. Deyà, *European Journal of Organic Chemistry*, Dec. 2007, **2007**, 5821–5825.
- (118) A. Garcia-Raso, F. M. Albertí, J. J. Fiol, A. Tasada, M. Barceló-Oliver, E. Molins, D. Escudero, A. Frontera, D. Quiñonero and P. M. Deyà, *Inorganic Chemistry*, Dec. 2007, **46**, 10724–10735.
- (119) P. Ballester, *Struct Bond*, 2008, **129**, 127–174.
- (120) A. García-Raso, F. M. Albertí, J. J. Fiol, A. Tasada, M. Barceló-Oliver, E. Molins, C. Estarellas, A. Frontera, D. Quiñonero and P. M. Deyà, *Crystal Growth & Design*, May 2009, **9**, 2363–2376.

-
- (121) S. E. Wheeler and K. N. Houk, *The Journal of Physical Chemistry A*, Aug. 2010, **114**, 8658–8664.
- (122) C. Estarellas, A. Bauzá, A. Frontera, D. Quiñonero and P. M. Deyà, *Physical Chemistry Chemical Physics*, Apr. 2011, **13**, 5696–5702.
- (123) J. C. A. Oliveira, J. Feldt, N. Galamba and R. A. Mata, *The Journal of Physical Chemistry A*, June 2012, **116**, 5464–5471.
- (124) D. Seebach, T. Kimmerlin, R. Šebesta, M. A. Campo and A. K. Beck, *Tetrahedron*, Aug. 2004, **60**, 7455–7506.
- (125) T. A. Martinek and F. Fülöp, *European Journal of Biochemistry*, Sept. 2003, **270**, 3657–3666.
- (126) D. Seebach, A. K. Beck and D. J. Bierbaum, *Chemistry & Biodiversity*, Aug. 2004, **1**, 1111–1239.
- (127) R. P. Cheng, S. H. Gellman and W. F. DeGrado, *Chemical Reviews*, Oct. 2001, **101**, 3219–3232.
- (128) D. Seebach, D. F. Hook and A. Glättli, *Biopolymers*, Jan. 2006, **84**, 23–37.
- (129) D. H. Appella, L. A. Christianson, D. A. Klein, D. R. Powell, X. Huang, J. J. Barchi and S. H. Gellman, *Nature*, May 1997, **387**, 381–384.
- (130) D. J. Hill, M. J. Mio, R. B. Prince, T. S. Hughes and J. S. Moore, *Chemical Reviews*, Dec. 2001, **101**, 3893–4012.
- (131) D. H. Appella, L. A. Christianson, I. L. Karle, D. R. Powell and S. H. Gellman, *Journal of the American Chemical Society*, Jan. 1996, **118**, 13071–13072.
- (132) D. Seebach, M. Overhand, F. N. M. Kühnle, B. Martinoni, L. Oberer, U. Hommel and H. Widmer, *Helvetica Chimica Acta*, June 1996, **79**, 913–941.
- (133) S. H. Choi, I. A. Guzei, L. C. Spencer and S. H. Gellman, *Journal of the American Chemical Society*, Oct. 2010, **132**, 13879–13885.
- (134) Y. J. Chung, B. R. Huck, L. A. Christianson, H. E. Stanger, S. Krauthäuser, D. R. Powell and S. H. Gellman, *Journal of the American Chemical Society*, May 2000, **122**, 3995–4004.
- (135) T. A. Martinek, G. K. Tóth, E. Vass, M. Hollósi and F. Fülöp, *Angewandte Chemie (International ed. in English)*, May 2002, **41**, 1718–21.

-
- (136) D. H. Appella, L. A. Christianson, D. A. Klein, M. R. Richards, D. R. Powell and S. H. Gellman, *Journal of the American Chemical Society*, Aug. 1999, **121**, 7574–7581.
- (137) G. Pohl, T. Beke-Somfai, I. G. Csizmadia and A. Perczel, *Amino Acids*, Aug. 2012, **43**, 735–749.
- (138) T. Beke, C. Somlai and A. Perczel, *Journal of Computational Chemistry*, Jan. 2006, **27**, 20–38.
- (139) T. Beke, I. G. Csizmadia and A. Perczel, *Journal of the American Chemical Society*, May 2006, **128**, 5158–5167.
- (140) Z. Lin, F. H. Hodel and W. F. van Gunsteren, *Helvetica Chimica Acta*, Apr. 2011, **94**, 597–610.
- (141) S. K. Pandey, G. F. Jogdand, J. C. A. Oliveira, R. A. Mata, P. R. Rajamohanan and C. V. Ramana, *Chemistry - A European Journal*, Nov. 2011, **17**, 12946–12954.
- (142) J. J. P. Stewart, *MOPAC2009*, Colorado Springs, CO, USA, 2008.
- (143) R. A. Kendall, T. H. Dunning and R. J. Harrison, *The Journal of Chemical Physics*, 1992, **96**, 6796–6806.
- (144) F. Neese, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, Jan. 2012, **2**, 73–78.
- (145) X. Zhu, P. Koenig, S. H. Gellman, A. Yethiraj and Q. Cui, *The Journal of Physical Chemistry B*, May 2008, **112**, 5439–5448.
- (146) B. Keller, Z. Gattin and W. F. van Gunsteren, *Proteins*, May 2010, **78**, 1677–1690.
- (147) F. Fülöp, T. A. Martinek and G. K. Tóth, *Chemical Society Reviews*, Apr. 2006, **35**, 323–334.
- (148) I. Rozas, *Physical Chemistry Chemical Physics*, June 2007, **9**, 2782–2790.
- (149) T. A. Martinek and F. Fülöp, *Chemical Society Reviews*, Jan. 2012, **41**, 687–702.
- (150) J. A. Kritzer, J. Tirado-Rives, S. A. Hart, J. D. Lear, W. L. Jorgensen and A. Schepartz, *Journal of the American Chemical Society*, Jan. 2005, **127**, 167–78.

-
- (151) A. Glättli, D. Seebach and W. F. van Gunsteren, *Helvetica Chimica Acta*, Oct. 2004, **87**, 2487–2506.
- (152) T. L. Raguse, J. R. Lai and S. H. Gellman, *Helvetica Chimica Acta*, Dec. 2002, **85**, 4154–4164.
- (153) E. Cubero, F. J. Luque, M. Orozco and J. Gao, *The Journal of Physical Chemistry B*, Feb. 2003, **107**, 1664–1671.
- (154) W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *The Journal of Chemical Physics*, 1983, **79**, 926–935.
- (155) M. W. Mahoney and W. L. Jorgensen, *The Journal of Chemical Physics*, 2000, **112**, 8910–8922.
- (156) U. W. Schmitt and G. A. Voth, *The Journal of Chemical Physics*, 1999, **111**, 9361–9381.
- (157) Y. Wu, H. L. Tepper and G. a. Voth, *The Journal of Chemical Physics*, Jan. 2006, **124**, 024503.
- (158) D. M. Benoit and D. C. Clary, *The Journal of Chemical Physics*, 2000, **113**, 5193–5202.
- (159) C. F. F. Karney, *Journal of Molecular Graphics & Modelling*, Jan. 2007, **25**, 595–604.
- (160) S. M. Thompson, K. E. Gubbins, J. P. R. B. Walton, R. A. R. Chantry and J. S. Rowlinson, *The Journal of Chemical Physics*, 1984, **81**, 530–542.
- (161) N. Galamba, *The Journal of Chemical Physics*, Sept. 2010, **133**, 124510.
- (162) T. J. Evans and T. N. Truong, *Journal of Computational Chemistry*, Nov. 1998, **19**, 1632–1638.
- (163) R. Sure and S. Grimme, *Journal of Computational Chemistry*, July 2013, **34**, 1672–1685.
- (164) J. W. Ponder, *TINKER—Software Tools for Molecular Design*, Saint Louis, 2013.

Curriculum Vitae

Personal information

Name:	João Carlos Agostinho de Oliveira
Address:	Hermann-Rein-Straße 8 37075 Göttingen Germany
Date of birth:	06.05.1982
Nationality:	Portuguese

Education

Since Apr. 2010	PhD student in Chemistry at the Computational Chemistry and Biochemistry group (CCB), Institute of Physical Chemistry, Georg-August-Universität Göttingen, Germany.
July 2007	Licenciatura in Chemistry at Faculty of Sciences University of Lisbon, Portugal.

Work experience

May 2010–Apr. 2014	Research assistant at the Computational Chemistry and Biochemistry group (CCB) at the Institute of Physical Chemistry from Georg-August-Universität Göttingen, under the supervision of Jun.-Prof. Dr. Ricardo Mata. (Germany)
Nov. 2008–Mar. 2010	Research assistant at the Center for Molecular Sciences and Materials (CCMM), Faculty of Sciences University of Lisbon and Atlantic University, under the supervision of Prof. Dr. Maria Filomena Camões, Prof. Dr. Sandra Félix and Dr. Cristina Oliveira. (Portugal)

- | | |
|----------------------|---|
| Sept. 2007–Aug. 2008 | Research assistant at the National Institute of Engineering, Technology and Innovation I. P. (INETI) on the Organic Synthesis Group of the Technology of Chemical Industries Department, under the supervision of Dr. Maria Alexandra Feijó Esteves. (Portugal) |
| Sep. 2004–Dec. 2006 | Internship within my Licenciatura degree at the Department of Chemistry and Biochemistry of the Faculty of Sciences University of Lisbon, under the supervision of Dr. Jorge Correia. (Portugal) |

Publications

- | | |
|----------------|--|
| September 2012 | J. M. Dieterich, J. C. A. Oliveira, and R. A. Mata, <i>Journal of Chemical Theory and Computation</i> , September 2012, 8, 3053–3060. |
| June 2012 | J. C. A. Oliveira, J. Feldt, N. Galamba, and R. A. Mata, <i>The Journal of Physical Chemistry A</i> , June 2012, 116, 5464–5471. |
| November 2011 | S. K. Pandey, G. F. Jogdand, J. C. A. Oliveira, R. A. Mata, P. R. Rajamohanan, and C. V. Ramana, <i>Chemistry – A European Journal</i> , November 2011, 17, 12946–12954. |
- The present publication also resulted in the inside cover of mentioned issue.

Communications

- | | |
|------|---|
| 2013 | João C. Oliveira, Jonas Feldt, Ricardo A. Mata, Development of a perturbative QM/MM Monte Carlo method for the study of molecules in solution. 112 th Bunsentagung (Annual German Conference on Physical Chemistry), 9-11 May 2013, Karlsruhe, Germany. (Poster) |
| 2012 | João C. A. Oliveira, Ricardo A. Mata, Conformational studies of β -peptides through the use of COSMO-LMP2. 4 th EuCheMS Chemistry Congress, 26-30 August 2012, Prague, Czech Republic. (Poster) |

- 2012 J. Oliveira, R. Mata, Theoretical study of amino acid-ion interactions through the use of LMP2 and DFT-SAPT. 5th Göttinger Chemie-Forum, 11 July 2012, Göttingen, Germany. (Poster)
- 2012 João Oliveira, Johannes M. Dieterich, Ricardo A. Mata, COSMO-LMP2 studies of biomolecules in solution. Molpro Workshop-Advanced Methods and Applications in Quantum Chemistry, 26-30 March 2012, Stuttgart, Germany. (Poster)
- 2011 João C. A. Oliveira, Ricardo A. Mata, Conformational studies of β -oligomers through the use of local correlation methods and a continuum solvation model. 9th Triennial Congress of the World Association of Theoretical and Computational Chemists (WATOC), 17-22 July 2011, Santiago de Compostela, Spain. (Poster)
- 2011 J. Oliveira, R. Mata, Conformational studies of β -oligomers through the use of local correlation methods and a continuum solvation model. 4th Göttinger Chemie-Forum, 8 June 2011, Göttingen, Germany. (Poster)
- 2011 João Oliveira, Johannes M. Dieterich, Ricardo A. Mata, Analytical LMP2 energy gradients with a continuum solvation model. 110th Bunsentagung (Annual German Conference on Physical Chemistry), 2-4 June 2011, Berlin, Germany. (Poster)
- 2010 Jonas Feldt, João Oliveira, Ricardo Mata, Exploring specific ion-amino acid Interactions through the use of local correlation methods. 46th Symposium on Theoretical Chemistry, 26-30 September 2010, Münster, Germany. (Poster)
- 2009 C. Oliveira, N. Martins, A. Caseiro, F. Mirante, C. Pio, M. Cerqueira, C. Alves, C. Oliveira, F. Camões, J. Oliveira, M. Matos, H. Silva, P. Cantinho and S. Félix, Atmospheric pollution in Lisbon urban atmosphere. European Geosciences Union-General Assembly 2009, 19-24 April 2009, Vienna, Austria. (Poster)

- 2008 J. C. Oliveira and J. P. Correia, Electroless Ni-P/Ru alloys for methanol oxidation in alkaline media. XV Meeting of the Portuguese Society of Electrochemistry, 3-6 September 2008, Faculty of Sciences University of Lisbon, Lisbon, Portugal. (Poster)
- 2008 M. Alexandra Esteves, João C. Oliveira, A. Margarida Guerreiro, Célia Santos, Carlos Baleizão, Chiral Vanadium(V) salen complexes immobilized on solids supports as heterogeneous catalysts for the Strecker reaction. XXI Iberoamerican Symposium of Catalysis, 22-27 June 2008, Benalmádena, Málaga, Spain. (Oral)
- 2008 M. Alexandra Esteves, João C. Oliveira, A. Margarida Guerreiro, Célia Santos, Carlos Baleizão, New Heterogeneous Chiral Vanadium(V) Salen Catalysts for the Strecker Reaction. XXI National Meeting of the Portuguese Society of Chemistry - Chemistry and Innovation, 11-13 June 2008, Engineering Faculty from Porto University, Porto, Portugal. (Poster)
- 2006 João C. Oliveira, Ruthenium Containing Electroless Nickel Alloys for Methanol Oxidation in Alkaline Media. Seminar given on the 25 May 2006, at the Chemistry and Biochemistry Center from Faculty of Sciences University of Lisbon, Lisbon, Portugal. (Seminar)
- 2005 J. C. Oliveira and J. P. Correia, Ruthenium Containing Electroless Nickel Alloys for Methanol Oxidation in Alkaline Media. XIII Meeting of the Portuguese Society of Electrochemistry, 6-9 July 2005, Beira Interior University, Covilhã, Portugal. (Poster)